

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 00:58:31 ; Search time 337.008 Seconds
(without alignments)
11326.533 Million cell updates/sec

Title: US-09-763-712A-1_COPY_1_1695

Perfect score: 1695

Sequence: 1 gtcacgaatctgcagcaaga.....cagtactgtcatctgcatta 1695

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2195239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SID22/gcgdata/geneseq/geneq-n-emb1/NA1980.DAT.*
- 2: /SID22/gcgdata/geneq/geneq-n-emb1/NA1981.DAT.*
- 3: /SID22/gcgdata/geneq/geneq-n-emb1/NA1982.DAT.*
- 4: /SID22/gcgdata/geneq/geneq-n-emb1/NA1983.DAT.*
- 5: /SID22/gcgdata/geneq/geneq-n-emb1/NA1984.DAT.*
- 6: /SID22/gcgdata/geneq/geneq-n-emb1/NA1985.DAT.*
- 7: /SID22/gcgdata/geneq/geneq-n-emb1/NA1986.DAT.*
- 8: /SID22/gcgdata/geneq/geneq-n-emb1/NA1987.DAT.*
- 9: /SID22/gcgdata/geneq/geneq-n-emb1/NA1988.DAT.*
- 10: /SID22/gcgdata/geneq/geneq-n-emb1/NA1989.DAT.*
- 11: /SID22/gcgdata/geneq/geneq-n-emb1/NA1990.DAT.*
- 12: /SID22/gcgdata/geneq/geneq-n-emb1/NA1991.DAT.*
- 13: /SID22/gcgdata/geneq/geneq-n-emb1/NA1992.DAT.*
- 14: /SID22/gcgdata/geneq/geneq-n-emb1/NA1993.DAT.*
- 15: /SID22/gcgdata/geneq/geneq-n-emb1/NA1994.DAT.*
- 16: /SID22/gcgdata/geneq/geneq-n-emb1/NA1995.DAT.*
- 17: /SID22/gcgdata/geneq/geneq-n-emb1/NA1996.DAT.*
- 18: /SID22/gcgdata/geneq/geneq-n-emb1/NA1997.DAT.*
- 19: /SID22/gcgdata/geneq/geneq-n-emb1/NA1998.DAT.*
- 20: /SID22/gcgdata/geneq/geneq-n-emb1/NA1999.DAT.*
- 21: /SID22/gcgdata/geneq/geneq-n-emb1/NA2000.DAT.*
- 22: /SID22/gcgdata/geneq/geneq-n-emb1/NA2001A.DAT.*
- 23: /SID22/gcgdata/geneq/geneq-n-emb1/NA2001B.DAT.*
- 24: /SID22/gcgdata/geneq/geneq-n-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|--------|---------------|--------|----|-----------------------------|
| 1 | 1695 | 100.0 | 2024 | 21 | AAA07697 Human collectin en |
| 2 | 1695 | 100.0 | 2628 | 22 | AAH43036 Nucleotide sequenc |
| 3 | 1693.4 | 99.9 | 2262 | 24 | ABA97932 Human scavenger re |
| 4 | 1693.4 | 99.9 | 2641 | 22 | AAC66903 Human EXMAD-14 cod |
| 5 | 1681.4 | 99.2 | 2929 | 24 | ABQ92072 Human polynucleoti |
| 6 | 1681.4 | 99.2 | 2930 | 20 | AAV55746 Human secreted pro |
| 7 | 1672.8 | 98.7 | 2181 | 22 | AAI60628 Human polynucleoti |
| 8 | 1661.8 | 98.0 | 2318 | 22 | AAI58842 Human polynucleoti |
| 9 | 1477.6 | 87.2 | 2005 | 24 | ABL95574 Human angiogenesis |

| | | | | | |
|----|--------|------|------|----|-----------------------------|
| 10 | 1477.6 | 87.2 | 2005 | 24 | ABL88085 Human PRO7223 cDNA |
| 11 | 1359 | 80.2 | 2637 | 22 | AAH43037 Nucleotide sequenc |
| 12 | 941 | 55.5 | 2256 | 22 | AAH43054 Nucleotide sequenc |
| 13 | 883.8 | 52.1 | 1521 | 23 | AAH71133 DNA encoding novel |
| 14 | 248.2 | 14.6 | 3685 | 20 | AAH27858 Human CSR3 protein |
| 15 | 248.2 | 14.6 | 3810 | 20 | AAH27856 DNA encoding novel |
| 16 | 169.4 | 10.0 | 493 | 23 | AAH71131 Human CSR1 protein |
| 17 | 150.2 | 8.9 | 873 | 23 | AAH71130 DNA encoding novel |
| 18 | 150.2 | 8.9 | 1062 | 23 | AAH67442 DNA encoding novel |
| 19 | 146.8 | 8.7 | 1329 | 23 | AAH71134 DNA encoding novel |
| 20 | 143 | 8.4 | 5041 | 22 | ABA17153 Human nervous syst |
| 21 | 135.6 | 8.0 | 1877 | 20 | AAH27857 Human CSR2 protein |
| 22 | 133.6 | 7.9 | 1707 | 22 | AAH44989 cDNA encoding nove |
| 23 | 127.2 | 7.5 | 986 | 22 | AAH89468 Human polynucleoti |
| 24 | 121.6 | 7.2 | 756 | 14 | AAH43034 Collagen-like poly |
| 25 | 121.6 | 7.2 | 756 | 17 | AAH16768 Collagen-like poly |
| 26 | 121.4 | 7.2 | 5676 | 15 | AAH64556 Human collagen (Ty |
| 27 | 121.4 | 7.2 | 8284 | 22 | AAH28526 Human breast cance |
| 28 | 121.4 | 7.2 | 8284 | 22 | AAH26553 Human breast cance |
| 29 | 121.4 | 7.2 | 8284 | 22 | AAH26600 Murine Col5a3 cDNA |
| 30 | 118.4 | 7.0 | 6109 | 23 | AAH78667 Collagen-like poly |
| 31 | 118 | 7.0 | 756 | 14 | AAH43032 Collagen-like poly |
| 32 | 118 | 7.0 | 756 | 17 | AAH16766 Human benign prost |
| 33 | 117 | 6.9 | 9287 | 24 | ABK64501 Bovine alpha1(I) c |
| 34 | 115.8 | 6.8 | 4748 | 22 | AAH06573 Mouse ischaemic co |
| 35 | 115 | 6.8 | 4270 | 24 | ABH99885 Human cDNA differe |
| 36 | 114.8 | 6.8 | 4816 | 24 | ABK84041 DNA encoding novel |
| 37 | 114.8 | 6.8 | 5058 | 23 | AAH69026 Human EST-derived |
| 38 | 114.8 | 6.8 | 6358 | 22 | AAH98286 DNA encoding novel |
| 39 | 114.8 | 6.8 | 6691 | 23 | AAH79806 DNA encoding novel |
| 40 | 114 | 6.7 | 5060 | 24 | ABK92176 Prostate cancer-as |
| 41 | 113.8 | 6.7 | 506 | 17 | AAH12198 pJG4-5-CDK-BP cDNA |
| 42 | 113.6 | 6.7 | 3181 | 19 | AAV59358 Nucleotide sequenc |
| 43 | 113.4 | 6.7 | 4821 | 23 | AAH68666 DNA encoding novel |
| 44 | 113.4 | 6.7 | 6200 | 23 | AAH78668 Human COL5A3 cDNA |
| 45 | 113.4 | 6.7 | 6368 | 24 | ABN59647 Novel human coding |

ALIGNMENTS

| | | | | | |
|----------|---|---|--|--|--|
| RESULT 1 | | | | | |
| AAA07697 | | | | | |
| ID | AAA07697 | standard; DNA; 2024 BP. | | | |
| XX | | | | | |
| AC | AAA07697; | | | | |
| XX | | | | | |
| DT | 20-JUN-2000 | (first entry) | | | |
| XX | | | | | |
| DE | Human collectin encoding DNA. | | | | |
| XX | | | | | |
| KW | Collectin; human; antibacterial; antiviral; ds. | | | | |
| XX | | | | | |
| OS | Homo sapiens. | | | | |
| XX | | | | | |
| FH | Key | Location/Qualifiers | | | |
| FT | CDS | 55..1698 | | | |
| FT | | /*tag= a | | | |
| FT | | /product= "collectin" | | | |
| FT | misc_feature | 1..738 | | | |
| FT | | /*tag= b | | | |
| FT | | /note= "fragment specifically claimed in claim 8" | | | |
| FT | misc_feature | 55..738 | | | |
| FT | | /*tag= c | | | |
| FT | | /note= "fragment specifically claimed in claim 8" | | | |
| FT | misc_feature | 79..738 | | | |
| FT | | /*tag= d | | | |
| FT | | /note= "fragment specifically claimed in claim 8" | | | |
| FT | misc_feature | 325..738 | | | |
| FT | | /*tag= e | | | |
| FT | | /note= "fragment specifically claimed in claim 8" | | | |
| FT | misc_feature | 358..738 | | | |
| FT | | /*tag= f | | | |

| | | |
|----|--|---|
| FT | misc_feature | /note= "fragment specifically claimed in claim 8" |
| FT | | 670...1695 |
| FT | | /*tag= g |
| FT | misc_feature | /note= "fragment specifically claimed in claim 5" |
| FT | | 685...738 |
| FT | | /*tag= h |
| FT | misc_feature | /note= "fragment specifically claimed in claim 7" |
| FT | | 730...738 |
| FT | | /*tag= i |
| FT | misc_feature | /note= "fragment specifically claimed in claim 7" |
| FT | | 739...1695 |
| FT | | /*tag= j |
| FT | misc_feature | /note= "fragment specifically claimed in claim 6" |
| FT | | 1696...2024 |
| FT | | /*tag= k |
| FT | | /note= "fragment specifically claimed in claim 9" |
| XX | | |
| XX | | |
| PN | WO200011161-A1. | |
| XX | | |
| XX | 02-MAR-2000. | |
| XX | | |
| XX | 24-AUG-1999; | 99WO-JP04552. |
| XX | | |
| PR | 24-AUG-1998; | 98JP-0237611. |
| XX | | |
| XX | (FUSO) FUSO PHARM IND LTD. | |
| XX | | |
| XX | Wakamiya N; | |
| PI | | |
| XX | | |
| DR | WPI; 2000-224596/19. | |
| DR | P-PSDB; AAY7985. | |
| XX | | |
| XX | | New collectin of human origin having antibacterial and antiviral activity, and gene encoding it useful for production of transgenic animals and of antibodies for screening potential drug molecules - |
| PT | | |
| PT | | |
| PT | | |
| XX | | |
| XX | Claim 5-9; Page 87-90; 106pp; Japanese. | |
| XX | | |
| CC | | The invention relates to polynucleotides encoding a new collectin of human origin. The collectin can be used as an antibacterial and antiviral agent and for screening potential drug molecules. The new collectin can be produced by standard recombinant methodology. The present sequence represents a DNA encoding the human collectin. |
| XX | | |
| XX | | |
| SQ | Sequence 2024 BP; 595 A; 524 C; 506 G; 399 T; 0 other; | |
| | Query Match | 100.0%; Score 1695; DB 21; Length 2024; |
| | Best Local Similarity | 100.0%; Pred. No. 0; |
| | Matches 1695; Conservative | 0; Mismatches 0; Indels 0; Gaps 0 |
| QY | 1 | GTCCAGAAATCTCAGCAAGATACCAAGCTGTCTCCAGGGCAATCTGCAGAACCAATCTAT 60 |
| Db | 1 | GTCCAGAAATCTCAGCAAGATACCAAGCTGTCTCCAGGGCAATCTGCAGAACCAATCTAT 60 |
| QY | 61 | TCTCATAATGTGGTCATCATGAACCTCAACACCTGAACCTGACCCAGGTGCAGCAGAGG 120 |
| Db | 61 | TCTCATAATGTGGTCATCATGAACCTCAACACCTGAACCTGACCCAGGTGCAGCAGAGG 120 |
| QY | 121 | AACCTCATCAGGAATCTGCAGGGTCTCTGGATGACACAAGCCAGGCTATCCAGCGAATC 180 |
| Db | 121 | AACCTCATCAGGAATCTGCAGGGTCTCTGGATGACACAAGCCAGGCTATCCAGCGAATC 180 |
| QY | 181 | AGAAGCGACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAGAGGACAGGATTGG 240 |
| Db | 181 | AGAAGCGACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAGAGGACAGGATTGG 240 |
| QY | 241 | CTGAAGGAGAAAGTGCAGAGCTTGACAGCGTGGCTGCCAACAACTCTGCGTTGGCCAAA 300 |
| Db | 241 | CTGAAGGAGAAAGTGCAGAGCTTGACAGCGTGGCTGCCAACAACTCTGCGTTGGCCAAA 300 |
| QY | 301 | GCCAAACAGCACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTTACAGGTCAGATG 360 |
| Db | 301 | GCCAAACAGCACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTTACAGGTCAGATG 360 |

| | | | |
|----------|---|--|------|
| Qy | 1441 | ATGGTAGGAGAGAGACCACTGGATCGGCCTCAGACTCAGAGCGTGAAATGAAATGG | 1500 |
| Db | 1441 | ATGGTAGGAGAGAGAGACCACTGGATCGGCCTCAGACTCAGAGCGTGAAATGAAATGG | 1500 |
| Qy | 1501 | AAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTGGAAGCTGGACAGCGGATAAC | 1560 |
| Db | 1501 | AAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTGGAAGCTGGACAGCGGATAAC | 1560 |
| Qy | 1561 | TGGGGTCATGGCCATGGCCAGGAGAACTGTGCTGGGTTGATTTATGCTGGGCAGTGG | 1620 |
| Db | 1561 | TGGGGTCATGGCCATGGCCAGGAGAACTGTGCTGGGTTGATTTATGCTGGGCAGTGG | 1620 |
| Qy | 1621 | AACGATTTCCAATGTGAAGACGTCGAATCACTTCATTCGCGAAAAAGACAGGAGACAGTA | 1680 |
| Db | 1621 | AACGATTTCCAATGTGAAGACGTCGAATCACTTCATTCGCGAAAAAGACAGGAGACAGTA | 1680 |
| Qy | 1681 | CTGTCATCTGCATTA | 1695 |
| Db | 1681 | CTGTCATCTGCATTA | 1695 |
| RESULT 2 | | | |
| AAH43036 | | | |
| ID | AAH43036 standard; DNA; 2628 BP. | | |
| XX | AAH43036; | | |
| XX | 15-OCT-2001 (first entry) | | |
| DT | Nucleotide sequence of a human scavenger receptor. | | |
| XX | Human; scavenger receptor; SRC1-P1; macrophage; basal immunity; | | |
| KW | arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty; | | |
| KW | low density lipoprotein; ss. | | |
| OS | Homo sapiens. | | |
| XX | Key | | |
| XX | Location/Qualifiers | | |
| FT | 74..2302 | | |
| FT | /*tag= a | | |
| FT | /product= "scavenger receptor" | | |
| XX | W0200159107-A1. | | |
| PN | 16-AUG-2001. | | |
| XX | 08-FEB-2001; 2001WO-JP00874. | | |
| XX | 14-FEB-2000; 2000JP-0035155. | | |
| PR | 10-OCT-2000; 2000JP-0309068. | | |
| XX | (FUSO) FUSO PHARM IND LTD. | | |
| PA | Wakamiya N; | | |
| PI | WPI; 2001-497076/54. | | |
| XX | P-PSDB; AAG63346. | | |
| DR | New Scavenger receptor proteins SRCL-P1 with collectin-like structure, | | |
| XX | useful for treatment and diagnosis of diseases associated with oxidized | | |
| PT | low-density lipoprotein accumulation - | | |
| PT | Claim 2; Page 79-84; 118pp; Japanese. | | |
| XX | The present sequence encodes a human scavenger receptor, designated | | |
| CC | SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They | | |
| CC | are useful in clarifying the functions of macrophages and basal | | |
| CC | immunity. They are also useful in the treatment, prevention, diagnosis | | |
| CC | and investigation of diseases such as arteriosclerosis, diabetic | | |
| CC | complications, bacterial infection and restenosis following angioplasty, | | |
| CC | which are associated with accumulation of oxidized low density | | |
| CC | lipoprotein and the binding of advanced glycation end-products into | | |
| CC | cells. | | |

Db 1565 GCGAAGGATCTAAAGGCTCCCAAGGCCCCCAAGGCTCCCGTGGTTCCCTCGGGAAGCC 1624
 QY 1021 GGCCCTCAGGCCCCAGTGGGACCCAGGCCCCCGGGCCCCACAGCAAGAGGACTC 1080
 Db 1625 GGCCCTCAGGCCCCAGTGGGACCCAGGCCCCCGGGCCCCACAGCAAGAGGACTC 1684
 QY 1081 CCGGCCCCAGGCCCCCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCTGGG 1140
 Db 1685 CCGGCCCCAGGCCCCCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCTGGG 1744
 QY 1141 GTGCTGGGACTCGGGACTCGGAGCTCGGAGCTTCCCTGGGGTACCAGGCATGCCAGCCCCAAG 1200
 Db 1745 GTGCTGGGACTCGGGACTCGGAGCTTCCCTGGGGTACCAGGCATGCCAGCCCCAAG 1804
 QY 1201 GGCCCCCCCCGCTCTCTGGCCCATCAGGAGCGTGGTGGCCCTGGCCCTGCAGATGAG 1260
 Db 1805 GGCCCCCCCCGCTCTCTGGCCCATCAGGAGCGTGGTGGCCCTGGCCCTGCAGATGAG 1864
 QY 1261 CCAACCCCGGACCGGAGGACAATGGCTGCGCCCTCACTGGGAAGAACTTCACAGACAAA 1320
 Db 1865 CCAACCCCGGACCGGAGGACAATGGCTGCGCCCTCACTGGGAAGAACTTCACAGACAAA 1924
 QY 1321 TGTACTATTTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGCTTTTCTGTGAAGAC 1380
 Db 1925 TGTACTATTTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGCTTTTCTGTGAAGAC 1984
 QY 1381 AAGCTTTCACATCTTCTTTTATAAACAACACTAGAGAGGACAGCAATGGATGATAAAAAACAG 1440
 Db 1985 AAGCTTTCACATCTTCTTTTATAAACAACACTAGAGAGGACAGCAATGGATGATAAAAAACAG 2044
 QY 1441 ATGGTAGGAGAGAGAGCCACTGGATGGCCCTCACAGACTCAGAGCGTGAATGAATGG 1500
 Db 2045 ATGGTAGGAGAGAGAGCCACTGGATGGCCCTCACAGACTCAGAGCGTGAATGAATGG 2104
 QY 1501 AAGTGGCTGGATGGGACATCTCAGACTACAAAAATTTGAAAGCTGGAGCGCGGATAAC 1560
 Db 2105 AAGTGGCTGGATGGGACATCTCAGACTACAAAAATTTGAAAGCTGGAGCGCGGATAAC 2164
 QY 1561 TGGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTGATTTATGCTGGGCGAGTGG 1620
 Db 2165 TGGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTGATTTATGCTGGGCGAGTGG 2224
 QY 1621 AAGCATTTCAATGTGAAGACGTCATAATCTTCATTTTGGAAAAAGACAGGGAGACAGTA 1680
 Db 2225 AAGCATTTCAATGTGAAGACGTCATAATCTTCATTTTGGAAAAAGACAGGGAGACAGTA 2284
 QY 1681 CTGTCTATCTGCATTA 1695
 Db 2285 CTGTCTATCTGCATTA 2299

RESULT 3
 ABA97932
 ID ABA97932 standard; cdna; 2262 BP.
 XX AC ABA97932;
 XX DT
 XX 25-APR-2002 (first entry)
 XX Human scavenger receptor-like protein encoding cdna SEQ ID NO 1.
 XX Human; scavenger; receptor; diagnosis; treatment; autoimmune disease;
 KW rheumatoid arthritis; gene; ss.
 XX OS Homo sapiens.
 XX Key Location/Qualifiers
 FT CDS 33..2262
 FT /tag= a
 FT /product= "scavenger receptor-like protein"
 XX JP2001340089-A.

PD 11-DEC-2001.
 XX 08-DEC-2000; 2000JP-0375066.
 XX 27-MAR-2000; 2000JP-0090772.
 XX (SHIO) SHIONOGI & CO LTD.
 XX WPI: 2002-144965/19.
 DR P-PSDB; ABB08642.
 XX New scavenger receptor-like protein for diagnosis, prevention and
 PT treatment of autoimmune disease, such as rheumatoid arthritis -
 PS Claim 3; Fig 1; 38pp; Japanese.
 CC The invention relates to a human scavenger receptor-like protein. The
 CC protein is useful as a target molecule for diagnosis, prevention and
 CC treatment of autoimmune diseases such as rheumatoid arthritis.
 XX Sequence 2262 BP; 670 A; 576 C; 581 G; 435 T; 0 other;
 Query Match 99.9%; Score 1693.4; DB 24; Length 2262;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1694; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GTCACGAATCTGCAGCAAGATACACGCGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
 Db 565 GTCACGAATCTGCAGCAAGATACACGCGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 624
 QY 61 TCTCATTAATGTGGTTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGACGAGAGG 120
 Db 625 TCTCATTAATGTGGTTCATCATGAACCTCAACAACCTGACCCAGGTGACGAGAGG 684
 QY 121 AACCTCATCAGAACTGCAGCGGTCTGTGGATGACACAGCCAGCTATCCAGCGAATC 180
 Db 685 AACCTCATCAGAACTGCAGCGGTCTGTGGATGACACAGCCAGCTATCCAGCGAATC 744
 QY 181 AAGAAGCACTTCAAAATCTGCAGCAGGTCTTCTTCAAGCCCAAGAGACACGCGATTGG 240
 Db 745 AAGAAGCACTTCAAAATCTGCAGCAGGTCTTCTTCAAGCCCAAGAGACACGCGATTGG 804
 QY 241 CTGAAGGAGAAAGTGCAGAGCTTGCAGCGTGGCTGCCAACAACCTCTCGTGGTGGCCAAA 300
 Db 805 CTGAAGGAGAAAGTGCAGAGCTTGCAGACACTGGCTGCCAACAACCTCTCGTGGTGGCCAAA 864
 QY 301 GCCACAACGACACCTCGGAGGATATGAACGACCTCAACTCATTCACAGTCAAGATG 360
 Db 865 GCCACAACGACACCTCGGAGGATATGAACGACCTCAACTCATTCACAGTCAAGATG 924
 QY 361 GAGAACATCACCCTATCTCTCAAGCCCAACGAGCAGAACTGAAAGACCTGCAGGACTTA 420
 Db 925 GAGAACATCACCCTATCTCTCAAGCCCAACGAGCAGAACTGAAAGACCTGCAGGACTTA 984
 QY 421 CACAAAGATGCAGAGAAATAGACAGCCATCAAGTTCAACCAACTGGAGGAAGCTTCCAG 480
 Db 985 CACAAAGATGCAGAGAAATAGACAGCCATCAAGTTCAACCAACTGGAGGAAGCTTCCAG 1044
 QY 481 CTCTTTGACAGCGATATTCTGACATCATTAGCAATATCAGTTACACAGCCCAACCTG 540
 Db 1045 CTCTTTGACAGCGATATTCTGACATCATTAGCAATATCAGTTACACAGCCCAACCTG 1104
 QY 541 CGGACGCTCACCAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAAA 600
 Db 1105 CGGACGCTCACCAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAAA 1164
 QY 601 CACACAGATGATCTGACCTCCTTGAATAATACCTGGCCCAACATCGGTTGGATTCGTT 660
 Db 1165 CACACAGATGATCTGACCTCCTTGAATAATACCTGGCCCAACATCGGTTGGATTCGTT 1224
 QY 661 TCTCTCAGGATGCAACAAGATTTGATGAGGTGAGGTTAGACACTGAAGTAGCCCACTTA 720
 Db 1225 TCTCTCAGGATGCAACAAGATTTGATGAGGTGAGGTTAGACACTGAAGTAGCCCACTTA 1284

Qy 721 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGCTCAGCTCATCAAGAAAT 780
 Db 1285 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGCTCAGCTCATCAAGAAAT 1344
 Qy 781 TTTACAATACATACAGGTCCACCGGGCCCGAGGGTCCAAAGAGTGACAGAGGATCCACAG 840
 Db 1345 TTTACAATACATACAGGTCCACCGGGCCCGAGGGTCCAAAGAGTGACAGAGGATCCACAG 1404
 Qy 841 GGACCCCTGGCCCACTGGCAACAAAGGACAGAAAGAGAGAGAGGGGAGGCTGGACCA 900
 Db 1405 GGACCCCTGGCCCACTGGCAACAAAGGACAGAAAGAGAGAGAGGGGAGGCTGGACCA 1464
 Qy 901 CCTGGCCCTGGGGGTGAGAGAGGCCAAATGGACAGCTGGTCCCGGAGAGCGTGGC 960
 Db 1465 CCTGGCCCTGGGGGTGAGAGAGGCCAAATGGACAGCTGGTCCCGGAGAGCGTGGC 1524
 Qy 961 GGCAAGGATCTAAGGCTCCACGGGCCCAAGGCTCCCGTGGTCCCGTGGGAGGCC 1020
 Db 1525 GGCAAGGATCTAAGGCTCCACGGGCCCAAGGCTCCCGTGGTCCCGTGGGAGGCC 1584
 Qy 1021 GGCCCTCAGGGCCCGAGTGGGACCCAGGCCCGCCCGGCCACCAAGCAAGAGGAGGACTC 1080
 Db 1585 GGCCCTCAGGGCCCGAGTGGGACCCAGGCCCGCCCGGCCACCAAGCAAGAGGAGGACTC 1644
 Qy 1081 CCGGCCCTCAGGGCCCTCCTGGCTTCCAGGAGACTTTCAGGGCACCGTTGGGAGGCTGG 1140
 Db 1645 CCGGCCCTCAGGGCCCTCCTGGCTTCCAGGAGACTTTCAGGGCACCGTTGGGAGGCTGG 1704
 Qy 1141 GTGCTGACCTCGGGGACTGCCAGGCTTGGCTGGGGTACCAAGCATGCCAGGCCCAAG 1200
 Db 1705 GTGCTGACCTCGGGGACTGCCAGGCTTGGCTGGGGTACCAAGCATGCCAGGCCCAAG 1764
 Qy 1201 GGCCCCCGCCCGCTCCTGGCCCAACAGAGGCGGTGGTGGCCCTGGCCCTGGCAGAAATGAG 1260
 Db 1765 GGCCCCCGCCCGCTCCTGGCCCAACAGAGGCGGTGGTGGCCCTGGCCCTGGCAGAAATGAG 1824
 Qy 1261 CCAACCCCGGACCGGAGGACAAATGGCTGCCCGCTCCTACTGGAAGAACTTCACAGACAAA 1320
 Db 1825 CCAACCCCGGACCGGAGGACAAATGGCTGCCCGCTCCTACTGGAAGAACTTCACAGACAAA 1884
 Qy 1321 TGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCTTTCTGTGTAAGAC 1380
 Db 1885 TGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCTTTCTGTGTAAGAC 1944
 Qy 1381 AAGCTTTCACATCTTGTGTTTTCATTAACACTAGAGAGGAAACAGCAATGGATGATAAAAAACAG 1440
 Db 1945 AAGCTTTCACATCTTGTGTTTTCATTAACACTAGAGAGGAAACAGCAATGGATGATAAAAAACAG 2004
 Qy 1441 ATGGTAGGAGAGAGAGCCACTGGATCGGCCTTCACAGACTCAGAGCGTGAAATGAATGG 1500
 Db 2005 ATGGTAGGAGAGAGAGCCACTGGATCGGCCTTCACAGACTCAGAGCGTGAAATGAATGG 2064
 Qy 1501 AAGTGGCTGGATGGACATCTCCAGACTTCCAGAAATTTGGAAGCTGGACCGCGGATAC 1560
 Db 2065 AAGTGGCTGGATGGACATCTCCAGACTTCCAGAAATTTGGAAGCTGGACCGCGGATAC 2124
 Qy 1561 TGGGTCTATGGCCATGGCCAGGAGAACTGTGCTGGTGTGATTTATGCTGGGCGAGTGG 1620
 Db 2125 TGGGTCTATGGCCATGGCCAGGAGAACTGTGCTGGTGTGATTTATGCTGGGCGAGTGG 2184
 Qy 1621 AACGATTTCCATGTGAAGAGCTCAATAAATTCATTTTCGAAAAAGACAGGGGAGACAGTA 1680
 Db 2185 AACGATTTCCATGTGAAGAGCTCAATAAATTCATTTTCGAAAAAGACAGGGGAGACAGTA 2244
 Qy 1681 CTGTCATCTGCATTA 1695
 Db 2245 CTGTCATCTGCATTA 2259

RESULT 4
 AAC66903
 ID AAC66903 standard; cDNA; 2641 BP.

XX AAC66903;
 AC 27-MAR-2001 (first entry)
 DT Human EXMAD-14 coding sequence SEQ ID NO: 39.
 DE Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
 XX inflammation; reproductive disorder; cardiovascular disorder;
 KW immune disorder; musculoskeletal disorder; developmental disorder;
 KW gastrointestinal disorder; cell proliferation disorder; ss.
 XX Homo sapiens.
 OS WO200068380-A2.
 PN 16-NOV-2000.
 XX 10-MAY-2000; 2000WO-US12811.
 PF 11-MAY-1999; 99US-0133643.
 XX 23-AUG-1999; 99US-0150409.
 PR (INCY-) INCYTE GENOMICS INC.
 XX Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DAM;
 PI Azimzai Y;
 XX WPI: 2001-007395/01.
 DR P-FSDB; AAB27236.
 XX Isolated polynucleotide encoding extracellular matrix or
 PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
 PT preventing disorders associated with expression of EXMAD such as
 PT proliferative, immune and genetic disorders -
 XX Claim 4; Page 121-122; 129pp; English.
 PS The present invention provides the protein and coding sequences for 25
 XX novel extracellular matrix and adhesion-associated proteins (EXMADS).
 CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
 CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
 CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
 CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
 CC useful in the prevention and treatment of cancers, cell proliferation,
 CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
 CC gastrointestinal disorders and inflammation.
 XX Sequence 2641 BP; 823 A; 628 C; 608 G; 582 T; 0 other;
 SQ Query Match 99.9%; Score 1693.4; DB 22; Length 2641;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1694; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 GTCAGCAATCTGCAGCAAGATACCAAGTCCAGGGCAATCTCAGAGCAATCTCAGAACCAATGTAT 60
 Db 243 GTCAGCAATCTGCAGCAAGATACCAAGTCCAGGGCAATCTCAGAGCAATCTCAGAACCAATGTAT 302
 Qy 61 TCTCATATATGGTTCATCATGNAACCTCAACACCTGAACCTGACCCAGGTGCAGCAGAGG 120
 Db 303 TCTCATATATGGTTCATCATGNAACCTCAACACCTGAACCTGACCCAGGTGCAGCAGAGG 362
 Qy 121 AACCTCATCAGCAATCTGCAGGGTCTGTGGATGACAAAGCCAGGCTATCCAGGCAATC 180
 Db 363 AACCTCATCAGCAATCTGCAGGGTCTGTGGATGACAAAGCCAGGCTATCCAGGCAATC 422
 Qy 181 AAGACGACTTTCAAATCTGCAGAGGTTTCTTTCAGCCAAAGACACGAGGATTGG 240
 Db 423 AAGACGACTTTCAAATCTGCAGAGGTTTCTTTCAGCCAAAGACACGAGGATTGG 482
 Qy 241 CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCTTGCAGAGCTTGCAGCAACACTCTGCGTTGGCCAAA 300
 Db 483 CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCTTGCAGAGCTTGCAGCAACACTCTGCGTTGGCCAAA 542

QY 301 GCCAACACGACACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCAGAGTCCAGATG 360
 |||||
 Db 543 GCCAACACGACACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCAGAGTCCAGATG 602
 |||||
 QY 361 GAGAACATCACCACTATCTCTCAAGCCAGGAGCAGAACTGAAAGACCTGCAGGACTTA 420
 |||||
 Db 603 GAGAACATCACCACTATCTCTCAAGCCAGGAGCAGAACTGAAAGACCTGCAGGACTTA 662
 |||||
 QY 421 CACAAAGATSCAGAGAAATAGAACAGCCATCAAGTCTCAACCACTGAGGAAAGCTTCCAG 480
 |||||
 Db 663 CACAAAGATSCAGAGAAATAGAACAGCCATCAAGTCTCAACCACTGAGGAAAGCTTCCAG 722
 |||||
 QY 481 CTCTTTGAGACGGATATTTGTGAACATCATTAGCAATATCAGTTTACAGACCCACCACTG 540
 |||||
 Db 723 CTCTTTGAGACGGATATTTGTGAACATCATTAGCAATATCAGTTTACAGACCCACCACTG 782
 |||||
 QY 541 CGAGCGTGCAGCAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAA 600
 |||||
 Db 783 CGAGCGTGCAGCAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAA 842
 |||||
 QY 601 CACACAGATGATCTGACCTCTTTGAATATACCTCGCCCAACATCCGTTTGGATTCTGTT 660
 |||||
 Db 843 CACACAGATGATCTGACCTCTTTGAATATACCTCGCCCAACATCCGTTTGGATTCTGTT 902
 |||||
 QY 661 TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTTAGACACTGAAGTAGCCAACTTA 720
 |||||
 Db 903 TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTTAGACACTGAAGTAGCCAACTTA 962
 |||||
 QY 721 TCAGTATATGGAAGAAATGAAGCTAGTAGACTCAAGATGGTCAAGTATCAAGAAT 780
 |||||
 Db 963 TCAGTATATGGAAGAAATGAAGCTAGTAGACTCAAGATGGTCAAGTATCAAGAAT 1022
 |||||
 QY 781 TTTTACAATACTACAAGTCTCACCGGCTCCAGGGTCCAAAGAGTGCACAGAGATCCCAAG 840
 |||||
 Db 1023 TTTTACAATACTACAAGTCTCACCGGCTCCAGGGTCCAAAGAGTGCACAGAGATCCCAAG 1082
 |||||
 QY 841 GGACCCCTCGGCCCACTGTGCAACAAAGGACACAGAAAGGAGAGAGGGGAGCGCTGGACCA 900
 |||||
 Db 1083 GGACCCCTCGGCCCACTGTGCAACAAAGGACACAGAAAGGAGAGAGGGGAGCGCTGGACCA 1142
 |||||
 QY 901 CTTGGCCCTCGGGTGCAGAGAGCCCAANTTGGACCACTGGTCTCCCTCCCGGAGAGCTGGC 960
 |||||
 Db 1143 CTTGGCCCTCGGGTGCAGAGAGCCCAANTTGGACCACTGGTCTCCCTCCCGGAGAGCTGGC 1202
 |||||
 QY 961 GGCAAGGATCTAAAGGCTCCCAAGGCCCCCAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCC 1020
 |||||
 Db 1203 GGCAAGGATCTAAAGGCTCCCAAGGCCCCCAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCC 1262
 |||||
 QY 1021 GGCCTCAGGGCCCCAGTGGGGACCCAGGCCCCCGGGCCCCCAGGCAAGAGGAGCTC 1080
 |||||
 Db 1263 GGCCTCAGGGCCCCAGTGGGGACCCAGGCCCCCGGGCCCCCAGGCAAGAGGAGCTC 1322
 |||||
 QY 1081 CCGGGCCCTCAGGGCCCTCTCTGCTTCCAGGGACTTCAGGGACCCGTTGGGGAGCTGGG 1140
 |||||
 Db 1323 CCGGGCCCTCAGGGCCCTCTCTGCTTCCAGGGACTTCAGGGACCCGTTGGGGAGCTGGG 1382
 |||||
 QY 1141 GTCCCTGGACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAAGGCATGCCAGGCCCAAG 1200
 |||||
 Db 1383 GTCCCTGGACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAAGGCATGCCAGGCCCAAG 1442
 |||||
 QY 1201 GGGCCCCCGGCCCTCTCTGGCCCATCAGGAGGGTGGTGGCCCTGGCCCTGCAGAAATGAG 1260
 |||||
 Db 1443 GGGCCCCCGGCCCTCTCTGGCCCATCAGGAGGGTGGTGGCCCTGGCCCTGCAGAAATGAG 1502
 |||||
 QY 1261 CCAACCCCGGCACCGGAGGACATGCTCCCGCTCACTGGAAGACTTCACAGACAAA 1320
 |||||
 Db 1503 CCAACCCCGGCACCGGAGGACATGCTCCCGCTCACTGGAAGACTTCACAGACAAA 1562
 |||||
 QY 1321 TGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAGAC 1380
 |||||
 Db 1563 TGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAGAC 1622
 |||||

QY 1381 AAGTCTTCACATCTTCTTTTTCATAAACTAGAGAGGAACAGCAATGGATAAAAAACAG 1440
 |||||
 Db 1623 AAGTCTTCACATCTTCTTTTTCATAAACTAGAGAGGAACAGCAATGGATAAAAAACAG 1682
 |||||
 QY 1441 ATGGTAGGAGAGAGAGCCACTGGGATCGGCTCACAGACTCAGAGCGTGAATGATGG 1500
 |||||
 Db 1683 ATGGTAGGAGAGAGAGCCACTGGGATCGGCTCACAGACTCAGAGCGTGAATGATGG 1742
 |||||
 QY 1501 AAGTGGCTGATGGGACATCTCCAGACTACAAAAATTGAAAGCTGCAGACCGCGATAAC 1560
 |||||
 Db 1743 AAGTGGCTGATGGGACATCTCCAGACTACAAAAATTGAAAGCTGCAGACCGCGATAAC 1802
 |||||
 QY 1561 TGGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTTGATTTATGCTGGGAGTGG 1620
 |||||
 Db 1803 TGGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTTGATTTATGCTGGGAGTGG 1862
 |||||
 QY 1621 AAGGATTTCCATGTGAAGACGTCATTAATCTTCATTTGCGAAAAAGACAGGAGACAGTA 1680
 |||||
 Db 1863 AAGGATTTCCATGTGAAGACGTCATTAATCTTCATTTGCGAAAAAGACAGGAGACAGTA 1922
 |||||
 QY 1681 CTGTCATCTGCATTA 1695
 |||||
 Db 1923 CTGTCATCTGCATTA 1937
 |||||
 RESULT 5
 ABQ92072
 ID ABQ92072 standard; cdna; 2929 BP.
 XX AC ABQ92072;
 XX DT 04-OCT-2002 (first entry)
 XX Human polynucleotide SEQ ID NO 69.
 DE Human; cytostatic; antirheumatic; antiarthritic; vulnerary; analgesic;
 KW antinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
 KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
 KW antulcer; fungicide; antidiabetic; antiasthmatic; antiallergic;
 KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
 KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
 KW stem cell; growth factor; nervous system disease; neuropathy;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
 KW osteoporosis; severe combined immunodeficiency; SCID; infection;
 KW multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
 XX Homo sapiens.
 OS US2002065394-A1.
 PN 30-MAY-2002.
 XX 22-DEC-2000; 2000US-0745763.
 PF 18-MAR-1998; 98US-0040963.
 PR (JACO/) JACOBS K.
 PA (MCCO/) MCCOY J M.
 PA (LAVA/) LAVALLIE E R.
 PA (COLL/) COLLINS-RACIE L A.
 PA (EVAN/) EVANS C.
 PA (MERB/) MERBERG D.
 PA (TREA/) TREACY M.
 PA (SPAU/) SPAULDING V.
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Spaulding V;
 XX WPI; 2002-582343/62.
 DR P-PSDB; ABP61859.
 XX Novel secreted or transmembrane protein and polynucleotide encoding the
 PT protein, useful for diagnosis and treatment of neurological disorders,

QY 1561 TGGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTTGATTTATCTGGGAGTGG 1620
 |||||
 Db 2157 TGGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTTGATTTATCTGGGAGTGG 2216
 |||||
 QY 1621 AAGGATTTCCATGTCGAACGCTCAATACTTCATTTGCGAAAAGACAGGGAGACAGTA 1680
 |||||
 Db 2217 AAGGATTTCCATGTCGAACGCTCAATACTTCATTTGCGAAAAGACAGGGAGACAGTA 2276
 |||||
 QY 1681 CTGTCACTGTCATTA 1695
 |||||
 Db 2277 CTGTCACTGTCATTA 2291

RESULT 6
 AAV55746
 ID AAV55746 standard; cDNA; 2930 BP.
 AC AAV55746;
 XX
 DT 23-MAR-1999 (first entry)
 XX
 DE Human secreted protein clone bv227_1 coding sequence.
 XX
 KW Secreted protein; human; nutritional supplements; immune stimulant;
 KW immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis;
 KW activin/inhibin; chemokinesis; haemostasis; thrombolysis;
 KW receptor/ligand activity; anti-inflammatory; tumour inhibitor;
 KW cadherin/tumour invasion suppressor; ds.
 XX
 OS Homo sapiens.

FH Key Location/Qualifiers
 FT CDS 67..693
 FT /*tag= a
 XX
 XX WO9855614-A2.
 XX
 PD 10-DEC-1998.
 XX
 PF 01-JUN-1998; 98WO-US11210.
 XX
 PR 29-MAY-1998; 98US-0087255.
 PR 04-JUN-1997; 97US-0868696.
 PR 04-JUN-1997; 97US-0868697.
 PR 04-JUN-1997; 97US-0868698.
 PR 04-JUN-1997; 97US-0868698.
 PR 04-JUN-1997; 97US-0868899.
 PR 04-JUN-1997; 97US-0868900.
 PR 04-JUN-1997; 97US-0869191.
 PR 04-JUN-1997; 97US-0869192.
 PR 04-JUN-1997; 97US-0869193.
 PR 04-JUN-1997; 97US-0869194.
 XX
 PA (GENY) GENETICS INST INC.

XX
 PI Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER;
 PI McCoy JW, Racie LA, Spaulding V, Treacy M;
 XX
 DR WPI: 1999-059912/05.
 DR P-PSDB; AAW73628.

XX New polynucleotides encoding secreted human proteins - derived from
 PT human foetal brain, adult brain, foetal kidney, adult ovary, adult
 PT retina, adult placenta or adult uterus cDNA libraries
 XX
 PS Claim 26; Page 87-88; 127pp; English.
 XX
 CC This sequence encodes a human secreted protein of the invention.
 CC This DNA sequence was isolated from a human adult brain cDNA
 CC library, and was designated clone bv227_1. The DNAs and proteins
 CC are predicted to have biological activities which would make them
 CC suitable for treating, preventing or ameliorating medical conditions in

CC humans and animals, although no supporting data is given. Suggested
 CC activities include nutritional sources or supplements, immune
 CC stimulating or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity. The DNAs
 CC are also stated to be useful for gene therapy. A host cell transfected
 CC with the DNA, or its subfragments and variants is useful for recombinant
 CC production of the human secreted protein clones.
 XX
 SQ Sequence 2930 BP; 902 A; 699 C; 692 G; 637 T; 0 other;

Query Match 99.2%; Score 1681.4; DB 20; Length 2930;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1693; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 1 GTCACGAATCTGCAGCAAGATACCAGCGTCTCCAGGGCAATCTGCAGAACCAATGTAT 60
 |||||
 Db 598 GTCACGAATCTGCAGCAAGATACCAGCGTCTCCAGGGCAATCTGCAGAACCAATGTAT 657
 |||||
 QY 61 TCTCATAATGTGTCTATCATGAACCTCAACAACCTGACCCAGGTGCAGCAGAGG 120
 |||||
 Db 658 TCTCATAATGTGTCTATCATGAACCTCAACAACCTGACCCAGGTGCAGCAGAGG 716
 |||||
 QY 121 AACCTCATCAGGAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGCTATCCAGCGAATC 180
 |||||
 Db 717 AACCTCATCAGGAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGCTATCCAGCGAATC 776
 |||||
 QY 181 AAGAAGGACCTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCCAAGAGCAGGATTGG 240
 |||||
 Db 777 AAGAAGGACCTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCCAAGAGCAGGATTGG 836
 |||||
 QY 241 CTGAAGGAGAAAAGTGCAGAGCTTGCAGAGCGTGGCTGCCAACAACCTGCGTTGGCCAAA 300
 |||||
 Db 837 CTGAAGGAGAAAAGTGCAGAGCTTGCAGAGCGTGGCTGCCAACAACCTGCGTTGGCCAAA 896
 |||||
 QY 301 GCCAACACGACACCCCTGGAGGATATGAACAGCCAGCTCAACTTATTCAGGTTCAGATG 360
 |||||
 Db 897 GCCAACACGACACCCCTGGAGGATATGAACAGCCAGCTCAACTTATTCAGGTTCAGATG 956
 |||||
 QY 361 GAGAACATCACCACCTATCTCTCAAGCCCAAGAGCAGCAACCTGAAAGACCTGCGAGGACTTA 420
 |||||
 Db 957 GAGAACATCACCACCTATCTCTCAAGCCCAAGAGCAGCAACCTGAAAGACCTGCGAGGACTTA 1016
 |||||
 QY 421 CACAAAGATGCAGAGAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACCTTCCAG 480
 |||||
 Db 1017 CACAAAGATGCAGAGAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACCTTCCAG 1076
 |||||
 QY 481 CTCCTTGGAGACGGATATTTGTGAACATCATTTAGCAATATCAGTTTACAGCCACACCTG 540
 |||||
 Db 1077 CTCCTTGGAGACGGATATTTGTGAACATCATTTAGCAATATCAGTTTACAGCCACACCTG 1136
 |||||
 QY 541 CGGACGCTGACCAGCAATCTAAATGAAGTTCAGGACCACTTGCACAGATACCCCTTACCAA 600
 |||||
 Db 1137 CGGACGCTGACCAGCAATCTAAATGAAGTTCAGGACCACTTGCACAGATACCCCTTACCAA 1196
 |||||
 QY 601 CACACAGATGATCTGACCTCCTTTGTAATAATACCTGGCCCAACATCCGTTTGGATTCTGTT 660
 |||||
 Db 1197 CACACAGATGATCTGACCTCCTTTGTAATAATACCTGGCCCAACATCCGTTTGGATTCTGTT 1256
 |||||
 QY 661 TCTCTCAGGATCAACAAGATTTGTAGGTGAGGTGTAGACACTGAAGTAGCCCAACTTA 720
 |||||
 Db 1257 TCTCTCAGGATCAACAAGATTTGTAGGTGAGGTGTAGACACTGAAGTAGCCCAACTTA 1316
 |||||
 QY 721 TCAGTGATTTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCATCAAGAAT 780
 |||||
 Db 1317 TCAGTGATTTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCATCAAGAAT 1376
 |||||
 QY 781 TTTTACAATACTACAGGTCACCCGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCCCAG 840
 |||||
 Db 1377 TTTTACAATACTACAGGTCACCCGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCCCAG 1436
 |||||

Db 192 TCGACGAGCTTTTCTTCAAGCCAAAGAGGACACCGATTGGCTGAAGGAGAAAGTGCAGA 251
QY 260 GCTTCGACAGCTGGCTGCCAAACACTCTGCGTTGGCCAAAGCCCAACACGACACCTCGG 319
Db 252 GCTTCGACAGCTGGCTGCCAAACACTCTGCGTTGGCCAAAGCCCAACACGACACCTCGG 311
QY 320 AGGATATGAACAGCCAGCTCAACTCATCTCAGAGTTCAGATGAGAGCAATCACCACTATCT 379
Db 312 AGGATATGAACAGCCAGCTCAACTCATCTCAGAGTTCAGATGAGAGCAATCACCACTATCT 371
QY 380 CTCACGCCAACGAGGAGACCTTGAAGACCTCGAGACCTTACACAAAGATGACAGAAATC 439
Db 372 CTCACGCCAACGAGGAGACCTTGAAGACCTCGAGACCTTACACAAAGATGACAGAAATC 431
QY 440 GAACAGCCATCAAGTTTCAACCACTGGAGAACGCTTCCAGCTCTTTGAGACGGATATTCG 499
Db 432 GAACAGCCATCAAGTTTCAACCACTGGAGAACGCTTCCAGCTCTTTGAGACGGATATTCG 491
QY 500 TGAACATCATTAAGCAATATCAGTTACAGCCCAACACCTGGGAGCTTGACCAAGATC 559
Db 492 TGAACATCATTAAGCAATATCAGTTACAGCCCAACACCTGGGAGCTTGACCAAGATC 551
QY 560 TAAATGAAGTCAAGGACCTTGCACAGATACCTTACCACAAACACACAGATGATCTGACCT 619
Db 552 TAAATGAAGTCAAGGACCTTGCACAGATACCTTACCACAAACACACAGATGATCTGACCT 611
QY 620 CCTTGAATTAATACCTTGGCCAAATCCGTTTGGATTCTCTCTCAGGATGCAACAAG 679
Db 612 CCTTGAATTAATACCTTGGCCAAATCCGTTTGGATTCTCTCTCAGGATGCAACAAG 671
QY 680 ATTGTATGAGTTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAGAAA 739
Db 672 ATTGTATGAGTTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAGAAA 731
QY 740 TGAAGCTAGTACTGACCTCAAGCATGTCAGCTCATCAAGAAATTTACAATACTACAGGTC 799
Db 732 TGAAGCTAGTACTGACCTCAAGCATGTCAGCTCATCAAGAAATTTACAATACTACAGGTC 791
QY 800 CACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGATCCACGAGGACCCCTGGCCCCAACATG 859
Db 792 CACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGATCCACGAGGACCCCTGGCCCCAACATG 851
QY 860 GCAACAGGACGACAGAAAGAGAGAGGGGAGCCTGGACCACTTGCCCTCGCGGTGAGA 919
Db 852 GCAACAGGACGACAGAAAGAGAGAGGGGAGCCTGGACCACTTGCCCTCGCGGTGAGA 911
QY 920 GAGGCCCAATTGACACAGCTGTCCTCCCGGAGAGCGTGGCGCAAGAGATCTAAAGGCT 979
Db 912 GAGGCCCAATTGACACAGCTGTCCTCCCGGAGAGCGTGGCGCAAGAGATCTAAAGGCT 971
QY 980 CCCAGGGCCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGGCCCCAGTG 1039
Db 972 CCCAGGGCCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGGCCCCAGTG 1031
QY 1040 GGGACCCAGCCCCCGGGCCCAACAGGCAAGAGGACTCCCGGCGCTCAGGGCCCTC 1099
Db 1032 GGGACCCAGCCCCCGGGCCCAACAGGCAAGAGGACTCCCGGCGCTCAGGGCCCTC 1091
QY 1100 CTGGCTTCCAGGACTTCAGGGCACCGTTGGGAGCCTGGGCTGCTTGACCTCGGGGAC 1159
Db 1092 CTGGCTTCCAGGACTTCAGGGCACCGTTGGGAGCCTGGGCTGCTTGACCTCGGGGAC 1151
QY 1160 TGCCAGGCTTGCTGGGTACAGGATGCCAGGCCCCAGGCCCCCGCCCTCCCTG 1219
Db 1152 TGCCAGGCTTGCTGGGTACAGGATGCCAGGCCCCAGGCCCCCGCCCTCCCTG 1211
QY 1220 GCGCATCAGAGGGTGTGGCCCTGGCCCTCGAGATCAGCAATCAGCCAGCCCGGAGG 1279
Db 1212 GCGCATCAGAGGGTGTGGCCCTGGCCCTCGAGATCAGCAATCAGCCAGCCCGGAGG 1271
QY 1280 ACAATGGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTG 1339
Db 1272 ACAATAGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTG 1331

QY 1340 AGAAGAAATTTTTCGAGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTGTT 1399
Db 1332 AGAAGAAATTTTTCGAGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTGTT 1391
QY 1400 TCATAAACACTAGAGAGGAAACAGCAATGATAAAAAACAAGATGGTAGGAGAGAGGCC 1459
Db 1392 TCATAAACACTAGAGAGGAAACAGCAATGATAAAAAACAAGATGGTAGGAGAGAGGCC 1451
QY 1460 ACTGGATCGGCTCAGACTCAGAGCTGAGAGCTGAAATGAATGGAAGTGGCTGGATGGACAT 1519
Db 1452 ACTGGATCGGCTCAGACTCAGAGCTGAGAGCTGAAATGAATGGAAGTGGCTGGATGGACAT 1511
QY 1520 CTCAGACTTACAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAATGCCCCATGGGC 1579
Db 1512 CTCAGACTTACAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAATGCCCCATGGGC 1571
QY 1580 CAGGAGAAAGCTGCTGGTTGATTATGCTGGGAGTGGACGATTTCCATGTTGAAG 1639
Db 1572 CAGGAGAAAGCTGCTGGTTGATTATGCTGGGAGTGGACGATTTCCATGTTGAAG 1631
QY 1640 AGCTCAATTAATTCATTTCGAAAAAGACAGGAGACAGTACTGTCTATCTGCATTA 1695
Db 1632 AGCTCAATTAATTCATTTCGAAAAAGACAGGAGACAGTACTGTCTATCTGCATTA 1687

RESULT 8

AA158842
ID AA158842 standard; cdna; 2318 BP.
XX AA158842;
AC AA158842;
XX
DT 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 1045.
DE
DE
XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW mekinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-052317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR P-PSDB: AAM39686.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
PS Claim 1; SEQ ID NO 1045; 10078pp; English.

The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAAM42213) with neurotropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed specification.

SQ Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other;

| | | | | |
|-----------------------|--------------|---------------|------------|--------------|
| Query Match | 98.0% | Score 1661.8; | DB 22; | Length 2318; |
| Best Local Similarity | 99.8% | Pred. No. 0; | | |
| Matches 1674; | Conservative | 0; | Mismatches | 2; Indels |
| | | | | 1; |

| | | | |
|----|-----|--|-----|
| Qy | 20 | ATACCGCGTCTCCAGGGCAATCTGCAGAACCAAAATGATTTCTCATTAATGTGGTCATCA | 79 |
| Db | 12 | ATACCGCGTCTCCAGGGCAATCTGCAGAACCAAAATGATTTCTCATTAATGTGGTCATCA | 71 |
| Qy | 80 | TGAACCTCAACAACCTGAACCTGACCCAGGTGCAGCAGAGGAACCTCATCAGGAATCTGC | 139 |
| Db | 72 | TGAACCTCAACAACCTGAACCTGACCCAGGTGCAGCAGAGGAACCTCATCAGGAATCTGC | 131 |
| Qy | 140 | AGCGTCTGTGGATGCACAAAGCCAGGCTATCCAGCGGAATCAAGAAGACACTTTCAAATC | 199 |
| Db | 132 | AGCGTCTGTGGATGCACAAAGCCAGGCTATCCAGCGGAATCAAGAAGACACTTTCAAATC | 191 |
| Qy | 200 | TGCAGCAGGTTTTTCTTCAAGCCAAAGAGGACACGGATGGCTGAAGAGAGAAAGTGCAGA | 259 |
| Db | 192 | TGCAGCAGGTTTTTCTTCAAGCCAAAGAGGACACGGATGGCTGAAGAGAGAAAGTGCAGA | 251 |
| Qy | 260 | GCTTGCAGACGCTGGCTGCCAAACAACTCTGCGTTGGCCAAAGCCAAACGACACCCCTGG | 319 |
| Db | 252 | GCTTGCAGACACTGGCTGCCAAACAACTCTGCGTTGGCCAAAGCCAAACGACACCCCTGG | 311 |
| Qy | 320 | AGGATATGAACAGCAGCTCAACTCATTTACAGAGTTCAGATGGAGAAACATCACCCTATCT | 379 |
| Db | 312 | AGGATATGAACAGCAGCTCAACTCATTTACAGAGTTCAGATGGAGAAACATCACCCTATCT | 371 |
| Qy | 380 | CTCAA - GCCAACGAGCAGAACCTGAAGACCTCGAGGACTTACACAAAGATGCAGAGAA | 438 |
| Db | 372 | CTCAAGGCCAACGAGCAGAACCTGAAGACCTCGAGGACTTACACAAAGATGCAGAGAA | 431 |
| Qy | 439 | AGAACGCCATCAAGTTCAACCAACTGGAGAACGCTTCCAGCTCTTTGAGACGGATATT | 498 |
| Db | 432 | AGAACGCCATCAAGTTCAACCAACTGGAGAACGCTTCCAGCTCTTTGAGACGGATATT | 491 |
| Qy | 499 | GTGAACATCATTAGCAATATCAGTTACACGCCACCACCTCGGACGCTGACCAGCAAT | 558 |
| Db | 492 | GTGAACATCATTAGCAATATCAGTTACACGCCACCACCTCGGACGCTGACCAGCAAT | 551 |
| Qy | 559 | CTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAAACACACAGATGATCTGACC | 618 |
| Db | 552 | CTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAAACACACAGATGATCTGACC | 611 |
| Qy | 619 | TCCTTGAATAATACCTGGCCAAACATCCGTTTGATTCGTCTCTCAGGATGCAACAA | 678 |
| Db | 612 | TCCTTGAATAATACCTGGCCAAACATCCGTTTGATTCGTCTCTCAGGATGCAACAA | 671 |
| Qy | 679 | GATTTGATGAGTCGAGGTTAGACACTGAACTAGCCAACTTATCAGTGATTTGGAAGA | 738 |
| Db | 672 | GATTTGATGAGTCGAGGTTAGACACTGAACTAGCCAACTTATCAGTGATTTGGAAGA | 731 |
| Qy | 739 | ATGAAGCTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAATTTTACAATACTACAAGGT | 798 |

RESULTS

ARL95574

ABL93374
ID ABL95574 standard: CDNA: 2005 BP.

AC ABL95574:

XX
XX
XXXXXX

DT 19-JUL-2002 (first entry)

Human angiogenesis related CDNA PRO7223 SEQ ID NO: 27:

| | |
|----|---|
| XX | Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer; |
| KW | atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder; |
| KW | cardiant; cyostatic; antiangiogenic; hypotensive; vulnerary; |
| KW | antiarteriosclerotic; gene; ss. |
| XX | |
| XX | Homo sapiens. |
| OS | |
| XX | WO200208284-A2. |
| XX | |
| XX | 31-JAN-2002. |
| PD | |
| XX | |
| PF | 09-JUL-2001; 2001WO-US21735. |
| XX | |
| XX | 20-JUL-2000; 2000US-219556P. |
| PR | 25-JUL-2000; 2000US-220624P. |
| PR | 25-JUL-2000; 2000US-220664P. |
| PR | 28-JUL-2000; 2000WO-US20710. |
| PR | 02-AUG-2000; 2000US-222695P. |
| PR | 17-AUG-2000; 2000US-0643657. |
| PR | 23-AUG-2000; 2000WO-US23522. |
| PR | 24-AUG-2000; 2000WO-US23328. |
| PR | 07-SEP-2000; 2000US-230378P. |
| PR | 15-SEP-2000; 2000US-000000P. |
| PR | 18-SEP-2000; 2000US-0664610. |
| PR | 18-SEP-2000; 2000US-0665350. |
| PR | 24-OCT-2000; 2000US-242922P. |
| PR | 08-NOV-2000; 2000US-0709238. |
| PR | 08-NOV-2000; 2000WO-US30952. |
| PR | 10-NOV-2000; 2000WO-US30873. |
| PR | 01-DEC-2000; 2000WO-US32678. |
| PR | 20-DEC-2000; 2000US-0747259. |
| PR | 20-DEC-2000; 2000WO-US34956. |
| PR | 22-JAN-2001; 2001US-0767609. |
| PR | 28-FEB-2001; 2001US-0796498. |
| PR | 28-FEB-2001; 2001WO-US06520. |
| PR | 01-MAR-2001; 2001WO-US06666. |
| PR | 09-MAR-2001; 2001US-0803706. |
| PR | 14-MAR-2001; 2001US-0808689. |
| PR | 22-MAR-2001; 2001US-0816744. |
| PR | 05-APR-2001; 2001US-0828366. |
| PR | 10-MAY-2001; 2001US-0854208. |
| PR | 10-MAY-2001; 2001US-0854280. |
| PR | 25-MAY-2001; 2001US-0866028. |
| PR | 25-MAY-2001; 2001US-0866034. |
| PR | 25-MAY-2001; 2001WO-US17092. |
| PR | 30-MAY-2001; 2001US-0870574. |
| PR | 30-MAY-2001; 2001WO-US17443. |
| PR | 01-JUN-2001; 2001WO-US17800. |
| PR | 20-JUN-2001; 2001WO-US19692. |
| PR | 28-JUN-2001; 2001WO-US00000. |
| XX | |
| PA | (GETH) GENENTECH INC. |
| PA | {BAKE/} BAKER K P. |
| PA | {FERR/} FERRARA N. |
| PA | {GERB/} GERBER H. |
| PA | {GERR/} GERRITSEN M E. |
| PA | {GODD/} GODDARD A. |
| PA | {GODO/} GODOWSKI P J. |
| PA | {GURN/} GURNEY A L. |
| PA | {HILL/} HILLAN K J. |
| PA | {MARS/} MARSTERS S A. |
| PA | {PANJ/} PAN J. |
| PA | {PAON/} PAONI N F. |
| PA | {STEP/} STEPHAN J F. |
| PA | {WATA/} WATANABE C K. |
| PA | {WILL/} WILLIAMS P M. |
| PA | {WOOD/} WOOD W I. |
| XX | |
| PI | Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A; |
| PI | Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF; |
| PI | Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W; |
| XX | |

| | | | | | |
|-----------------------|------|---|---------------|-----------|--------------|
| Query Match | | 87.2%; | Score 1477.6; | DB 24; | Length 2005; |
| Best Local Similarity | | 99.7%; | Pred. No. 0; | | |
| Matches 1491; | | Conservative 0; | Mismatches 4; | Indels 1; | Gaps 1; |
| QY | 200 | TGCAGCAGGTTTCTTCAAGCCAAAGACACGAGTGGTGAAGAGAAAGTGCAGA | 259 | | |
| Db | 9 | TCGGCAGGTTTTCTTCAAGCCAAAGACACGAGTGGTGAAGAGAAAGTGCAGA | 68 | | |
| QY | 260 | GCTTGACAGCGCTGGCTGCCAACACTCTGCGTTGGCCAAAGCCCAACAGCACCCCTCG | 319 | | |
| Db | 69 | GCTTGACAGACATGGCTGCCAACACTCTGCGTTGGCCAAAGCCCAACAGCACCCCTCG | 128 | | |
| QY | 320 | AGGATATGAACGCCAGCTCAACTCTATTCACAGGTCAGATGGAGAACATCACCCTATCT | 188 | | |
| Db | 129 | AGGATATGAACGCCAGCTCAACTCTATTCACAGGTCAGATGGAGAACATCACCCTATCT | 188 | | |
| QY | 380 | CTCAAGCCAAAGCAGAGAACCTTGAAGACCTCTGAGAGCTTACACAAAGATGACAGAAAT | 439 | | |
| Db | 189 | CTCAAGCCAAAGCAGAGAACCTTGAAGACCTCTGAGAGCTTACACAAAGATGACAGAAAT | 248 | | |
| QY | 440 | GAACAGCCATCAAGTTCAACCACTGAGGAGAGCGTTCCAGCTCTTTGAGAGCGATATTG | 499 | | |
| Db | 249 | GAACAGCCATCAAGTTCAACCACTGAGGAGAGCGTTCCAGCTCTTTGAGAGCGATATTG | 308 | | |
| QY | 500 | TGAACATCATATTAGCAATATCAGTTACACAGCCACACCTCGGGAGCGTGAACAGCAATC | 559 | | |
| Db | 309 | TGAACATCATATTAGCAATATCAGTTACACAGCCACACCTCGGGAGCGTGAACAGCAATC | 368 | | |
| QY | 560 | TAAATCAAGTCAGGACCACTTGCACAGATACCTTTACCAAAACACACAGATGATCTGACCT | 619 | | |
| Db | 369 | TAAATCAAGTCAGGACCACTTGCACAGATACCTTTACCAAAACACACAGATGATCTGACCT | 427 | | |
| QY | 620 | CCTTGAATTAATACCTTGGCCAAACATCCGTTTGGATTCTCTCTCAGGATGCAACAAG | 679 | | |
| Db | 428 | CCTTGAATTAATACCTTGGCCAAACATCCGTTTGGATTCTCTCTCAGGATGCAACAAG | 487 | | |
| QY | 680 | ATTGTGATGAGTGGAGGTTAGACACTGAAAGTAGCCAACTTATCAGTGATTATGGAAGAAA | 739 | | |
| Db | 488 | ATTGTGATGAGTGGAGGTTAGACACTGAAAGTAGCCAACTTATCAGTGATTATGGAAGAAA | 547 | | |
| QY | 740 | TGAAGCTAGTACCTCAACAGCATGGTCAGTCTATCAAGATTTTACAATACTACAAGTTC | 799 | | |
| Db | 548 | TGAAGCTAGTACCTCAACAGCATGGTCAGTCTATCAAGATTTTACAATACTACAAGTTC | 607 | | |
| QY | 800 | CACCGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCCAAGGACCCCTCGGCCCAACTG | 859 | | |
| Db | 608 | CACCGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCCAAGGACCCCTCGGCCCAACTG | 667 | | |
| QY | 860 | GCNACAGGACAGAAAGAGAGAGAGGGGAGCGCTGGACCACTTGGCCCTCGGGGTGAGA | 919 | | |
| Db | 668 | GCNACAGGACAGAAAGAGAGAGAGGGGAGCGCTGGACCACTTGGCCCTCGGGGTGAGA | 727 | | |
| QY | 920 | GAGGCCAATTGGACCACTGGTCCCGCCGAGAGCGTGGCGCAAGAGATCTAAGGCT | 979 | | |
| Db | 728 | GAGGCCAATTGGACCACTGGTCCCGCCGAGAGCGTGGCGCAAGAGATCTAAGGCT | 787 | | |
| QY | 980 | CCAGGGCCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGGCCCCAGTG | 1039 | | |
| Db | 788 | CCAGGGCCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGGCCCCAGTG | 847 | | |
| QY | 1040 | GGGACCCAGGCCCCCGGGCCACACAGGCAAAAGAGGACTCCCGGCGCTCAGGGCCCTC | 1099 | | |
| Db | 848 | GGGACCCAGGCCCCCGGGCCACACAGGCAAAAGAGGACTCCCGGCGCTCAGGGCCCTC | 907 | | |
| QY | 1100 | CTGGCTTCAGGAGCTTCAAGGCACCGTTGGGAGCGCTGGGCTGGAGCTCGGGGAC | 1159 | | |
| Db | 908 | CTGGCTTCAGGAGCTTCAAGGCACCGTTGGGAGCGCTGGGCTGGAGCTCGGGGAC | 967 | | |
| QY | 1160 | TGCCAGGCTTGGCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCGGCTCCCTG | 1219 | | |
| Db | 968 | TGCCAGGCTTGGCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCGGCTCCCTG | 1027 | | |

| | | | |
|----|------|---|------|
| QY | 1220 | GCCCATCAGAGCGGTGGTGGCCCTGCGCCCTGCAGAAATGAGCCAAACCCGGCACCGGAGG | 1279 |
| Db | 1028 | GCCCATCAGAGCGGTGGTGGCCCTGCGCCCTGCAGAAATGAGCCAAACCCGGCACCGGAGG | 1087 |
| QY | 1280 | ACAATGGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTCAGTTG | 1339 |
| Db | 1088 | ACAATAGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTCAGTTG | 1147 |
| QY | 1340 | AGAAAGAAATTTTGGAGGATGCAAAAGCTTTCTGTGAAGACAAAGTCTTCACATCTTTGTTT | 1399 |
| Db | 1148 | AGAAAGAAATTTTGGAGGATGCAAAAGCTTTCTGTGAAGACAAAGTCTTCACATCTTTGTTT | 1207 |
| QY | 1400 | TCATAAACACTAGAGAGAACAGCAATGGATAAAAAACAGATGTTAGGAGAGAGAGCC | 1459 |
| Db | 1208 | TCATAAACACTAGAGAGAACAGCAATGGATAAAAAACAGATGTTAGGAGAGAGAGCC | 1267 |
| QY | 1460 | ACTGGATCGGCTCAGAGACTCAGAGCGTGAAGTGAATGGAAGTGGCTGGATGGACAT | 1519 |
| Db | 1268 | ACTGGATCGGCTCAGAGACTCAGAGCGTGAAGTGAATGGAAGTGGCTGGATGGACAT | 1327 |
| QY | 1520 | CTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGATAACTGGGGTCTATGGCCATGGGC | 1579 |
| Db | 1328 | CTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGATAACTGGGGTCTATGGCCATGGGC | 1387 |
| QY | 1580 | CAGGAGAAAGACTGTGCTGGTTGATTTATGCTGGGCAAGTGGAAACAGATTTCCAATGTGAAG | 1639 |
| Db | 1388 | CAGGAGAAAGACTGTGCTGGTTGATTTATGCTGGGCAAGTGGAAACAGATTTCCAATGTGAAG | 1447 |
| QY | 1640 | AGCTCAATAACTTCATTTCCGAAAAAGACAGGAGACAGTACTGTGCATCTGCATTA | 1695 |
| Db | 1448 | AGCTCAATAACTTCATTTCCGAAAAAGACAGGAGACAGTACTGTGCATCTGCATTA | 1503 |

| | |
|-----------|---|
| RESULT 11 | |
| AAH43037 | |
| ID | AAH43037 standard; cDNA; 2637 BP. |
| XX | AAH43037; |
| XX | AC |
| DT | 15-OCT-2001 (first entry) |
| XX | Nucleotide sequence of a human scavenger receptor. |
| DE | Human; scavenger receptor; SRCL-P1; macrophage; basal immunity; |
| KW | arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty; |
| KW | low density lipoprotein; ss. |
| XX | Homo sapiens. |
| XX | Key |
| FT | CDS |
| FT | 92..2320 |
| FT | /*tag= a |
| FT | /product= "scavenger receptor" |
| XX | WO200159107-A1. |
| PN | 16-AUG-2001. |
| PD | 08-FEB-2001; 2001WO-JP00874. |
| XX | 14-FEB-2000; 2000JP-0035155. |
| PR | 10-OCT-2000; 2000JP-0309068. |
| XX | (FUSO) FUSO PHARM IND LTD. |
| XX | Wakamiya N; |
| PI | WPI; 2001-497076/54. |
| DR | P-PSDB; AAG63347. |
| XX | New Scavenger receptor proteins SRCL-P1 with collectin-like structure, |
| PT | useful for treatment and diagnosis of diseases associated with oxidized |
| PT | low-density lipoprotein accumulation |

XX PS Claim 6; Page 88-93; 118pp; Japanese.

XX CC The present sequence encodes a human scavenger receptor, designated

CC SRCL-p1. The SRCL-p1 polypeptide has a collectin-like structure. They

CC are useful in clarifying the functions of macrophages and basal

CC immunity. They are also useful in the treatment, prevention, diagnosis

CC and investigation of diseases such as arteriosclerosis, diabetic

CC complications, bacterial infection and restenosis following angioplasty,

CC which are associated with accumulation of oxidized low density

CC lipoprotein and the binding of advanced glycation end-products into

CC cells.

XX SQ Sequence 2637 BP; 829 A; 617 C; 635 G; 556 T; 0 other;

Query Match 80.2%; Score 1359; DB 22; Length 2637;

Best Local Similarity 87.6%; Pred. No. 0;

Matches 1485; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 1 GTCAGCAATCTGCAGCAAGATACCAAGCTGCTCCAGGGCAATCTGCAGAACCAATGTAT 60

DB 623 GTCAAAATCTGCAAAAGATACTAGTGTCTCCAGGGCAATCTGCAGAGCCAAATGTAT 682

QY 61 TCTCATATGTGGTCAATCAATCAACCTCAACCTGAACCTCAACCTCAACCTCAACCT 120

DB 683 TCTCAGAGCGTGGTATCATATGAACCTCAACAACTCAACCTCAACCTCAACCTCAACCT 742

QY 121 AACCTCATCAACCAATCTGCAGCGGTCTGTGGATGACAAAGCCAGGCTATCAGCGAATC 180

DB 743 AACCTTAATCAATCTGCAGCGGTCTGTGGATGACAAAGCCAGGCTATCAGCGAATC 802

QY 181 AAGAACGACTTTCAAAATCTGCAGCGGTCTGTGGATGACAAAGCCAGGCTATCAGCGAAT 240

DB 803 AAGAATGATTTCCAAATCTGCAGCGGTCTGTGGATGACAAAGCCAGGCTATCAGCGAAT 862

QY 241 CTGAGGAGAAAGTGCAGAGCTTGACAGCGTGGCTGCCAACAATCTCGCTGGCCCAA 300

DB 863 CTAAAGGAAAAAGTACAGAGCTTGACAGCGTGGCTGCCAACAATCTCGCTGGCCCAA 922

QY 301 GCCAACAAGCACCTTGAGGATATGAACAGCGAGTCAACCTCAATTCACAGGTCAGATG 360

DB 923 GCCAACAATGACACCTTAGAGGATATGAATAGCCAGCTCAGCTCAATTCACAGGTCAGATG 982

QY 361 GAGAACATACCACTATCTCTCAAGCCCAAGCAGCAGCACTGAAAGACCTGCGAGCACTTA 420

DB 983 GACAACTTACCACTATCTCTCAAGCCCAAGCAGCAGCACTGAAAGACCTTTCAGGACTTA 1042

QY 421 CACAAGATGACAGAAATAGACAGCGATCAAGTTCAACCACTGAGGAACGCTTCCAG 480

DB 1043 CACAAGGATACAGAAATAGACAGCTGTCAAGTTTCAGGCAACTCTGCGCTGGCCCAA 1102

QY 481 CTCTTTGAGACGGATATGTGAACATCATTTAGCAATATCAGTTACACAGCCCAACCTG 540

DB 1103 GTCTTTGAGACAGATATGTGAACATCATTTAGCAATATCAGTTACACAGCCCAACCTG 1162

QY 541 CGGACCTGACAGCAATCTAAATGAAGTCAGGACCACTTGACAGATACCTTACCAA 600

DB 1163 AGGACACTGACAGCAATCTGAATGATGTAGGACCACTGACAGACACCTTGACCAGA 1222

QY 601 CACACAGATGATCTGACCTCTTGAATATACCTTGGCCCAACATCGGTTGGATCTGTT 660

DB 1223 CACAGGGATGACCTGACCTCTTGAATATACCACTAGTCAACATCGGCTTGGATCTATT 1282

QY 661 TCTCTCAGGATCAACAGATTTGATGAGTGTAGGCTAGACACTCAAGTACCCCACTTA 720

DB 1283 TCTCTCAGGATCAACAGCAATGATGAGTGTAGGCTCAAGTGTAGACACTCAAGTACCCCACTTA 1342

QY 721 TCAGTGATTTGGAAGAAATGAAGCTAGTACTCAAGCATGGTCAGCTCATCAAGAT 780

DB 1343 TCAGTGATTTGGAAGAGATGAAGCTAGTACTCAAGCATGGTCAGCTCATCAAGAT 1402

QY 781 TTTACAATACTACAGGTCACCGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCCG 840

||||| ||| |||||||||| ||| |||||| |||||| |||||| |||||| |||||| |||

DB 1403 TTTACCATCTTACAAGGTCTCTCTGGCCCCAGAGGTCTCCAAAAGGTGACAGAGGATCTCAG 1462

QY 841 GGACCCCTTGGCCCCCAACTGGCAACAGGACAGAAAGAGAGAGAGGGGAGCCTGACCA 900

DB 1463 GGACCACTTGGTCCCAACTGGCAACAGGACAGAAAGAGAGAGAGGGGAGCCTGCTCCA 1522

QY 901 CTTGGCCCTTGGGGGTGAGAGAGCCCAATTTGGACAGCTTGGTCCCCCCCCGGAGAGCCTGGC 960

DB 1523 CTTGGCCCTTGGGGGTGAGAGAGCCCAATTTGGACAGCTTGGTCCCCCCCCGGAGAGCCTGGC 1582

QY 961 GGCAGAGGATCTTAAGGCTCCAGAGGCCCAAGGCTCCCGTGGTTCCTCTGGAGGCC 1020

DB 1583 AGCAAGGATCCAAAGGCTCACAGGCTCCCAAGGATCTCTGGGTCCCCAGGGAAGCCT 1642

QY 1021 GGCCCTCAGGGCCCAAGTGGGACCCAGGCCCCCGGGCCCAAGGCAAGAGGAGCCTC 1080

DB 1643 GGCCCTCAGGACCTAGTGGGACCCAGGACCCAGGCTCCAGGCAAGGATGAGCTC 1702

QY 1081 CCGGCGCTCAGGGCCCTCTCTGGTTCAGGACCTTTCAGGCAACCTTGGGAGCCTGGG 1140

DB 1703 CTTGGCCCTCAGGGCCCTCTCTGGTTCAGGACCTTTCAGGCAACCTTGGGAGCCTGGG 1762

QY 1141 GTGCTTGGACCTCGGGGCTGCGGCTTGGTGGGTACCGGCTCCAGGCCCCCAAG 1200

DB 1763 GTACCTGGACCTCGGGGCTTGGCAGGCTTGGCAGGCTGCTGGGCCCCAAG 1822

QY 1201 GGCCCCCGGGCCCTCTCTGGCCATCAGGAGCGGTGGTGGCCCTGCGCCTGAGAAATGAG 1260

DB 1823 GGACCACTTGGCCCTCCAGGCCCCCTCAGGACCAATGGAGCCCTTGGCTCTGAGAAATGA 1882

QY 1261 CCAACCCCGGACCCGGAGGACCAATGGCTGCGCCCTCTCTGGAAGAACTTTCAGAGCAAA 1320

DB 1883 CCAACCCCGAGCATCAGAGGTCAACGGATCTCGCCCTCACTGGAAGAACTTTCAGAGTAA 1942

QY 1321 TGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTCTGTGAAGAC 1380

DB 1943 TGCTACTATTTTTCATTTGAAAAGAAATTTTGAAGATGCTAAGCTTTCTGTGAAGAC 2002

QY 1381 AAGTCTTCACTCTGTTTTCATTAACACTAGAGAGAAACAGCAATGGATAAAAAACAG 1440

DB 2003 AATCTTCCATCTCTGTTTTCATTAACACTAGAGAGAAACAGCAATGGATAAAAAGCAT 2062

QY 1441 ATGGTAGGAGAGAGAGCCACTGGATCGCCCTCAGAGCTCAGAGCGTGAATGAATGG 1500

DB 2063 ACCGTGGGAGAGAAAGCCATTTGGATCGCCCTCAGAGCTCAGAGCAAGGAGCAATGG 2122

QY 1501 AAGTGGCTGGATGGGACATCTCCAGACTACAAAATTTGAAAAGCTGGAGAGCCGGATAC 1560

DB 2123 AAGTGGCTAGAGCGGCTCACTCTGTTGATTACAAAAGCTGGAAAAGCTGGACCAACAGATAAC 2182

QY 1561 TGGGGTCAATGGCCATGGGCGCAGGAGAAAGACTGTGTGGTGTGATTTATGCTGGCAGTGG 1620

DB 2183 TGGGCGAGTGGCCATGGGCGCAGGAGAAAGACTGTGTGGTGTGATTTATGCTGGCAGTGG 2242

QY 1621 AACGATTTCCAATGTGAAGACGCTCAATAACTTCAATTTGCGAAAAGAGAGAGGAGACAGTA 1680

DB 2243 AATGACTTCCAGTGTGATGAATCAATAACTTCAATTTGTGAGAAAGGAGGAGGAGCA 2302

QY 1681 CTGTCTATCTGCATTA 1695

DB 2303 CCATCATCATATTA 2317

RESULT 12

AAH43054

ID AAH43054 standard; DNA; 2256 BP.

XX

AC

AAH43054;

XX

15-OCT-2001 (first entry)

XX

DE Nucleotide sequence of a human scavenger receptor.

XX

KW Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;
 KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
 KW low density lipoprotein; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
 FH CDS 74..1936
 FT /*tag= a
 FT /product= "scavenger receptor"

XX W0200159107-A1.

XX 16-AUG-2001.

XX 08-FEB-2001; 2001WO-JP00874.

XX 14-FEB-2000; 2000JP-0035155.

XX 10-OCT-2000; 2000JP-0309068.

XX (FUSO) FUSO PHARM IND LTD.

XX Wakamiya N;

XX WPI; 2001-497076/54.

XX P-PSDB; AAG63350.

XX New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
 PT useful for treatment and diagnosis of diseases associated with oxidized
 PT low-density lipoprotein accumulation

XX Claim 4; Page 105-109; 118pp; Japanese.

XX The present sequence encodes a human scavenger receptor, designated
 CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
 CC are useful in clarifying the functions of macrophages and basal
 CC immunity. They are also useful in the treatment, prevention, diagnosis
 CC and investigation of diseases such as arteriosclerosis, diabetic
 CC complications, bacterial infection and restenosis following angioplasty,
 CC which are associated with accumulation of oxidized low density
 CC lipoprotein and the binding of advanced glycation end-products into
 CC cells.

XX Sequence 2256 BP; 721 A; 524 C; 528 G; 483 T; 0 other;

XX Query Match 55.5%; Score 941; DB 22; Length 2256;

XX Best Local Similarity 78.1%; Pred. No. 5.1e-229;

XX Matches 1323; Conservative 0; Mismatches 0; Indels 372; Gaps 1;

QY 1 GTCACGAATCTGCAGCAAGATACCAAGCTGCTCCAGGGCAATCTGCAGAACCAATGTAT 60

DB 605 GTCACGAATCTGCAGCAAGATACCAAGCTGCTCCAGGGCAATCTGCAGAACCAATGTAT 654

QY 61 TCTCATTAATGTGTCATCATGACCTCACACCTCAACCTGACCCAGCTGCAGCAGAGG 120

DB 665 TCTCATTAATGTGTCATCATGACCTCACACCTCAACCTGACCCAGCTGCAGCAGAGG 724

QY 121 AACCTCATACGAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGCAATC 180

DB 725 AACCTCATACGAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGCAATC 784

QY 181 AGAAGCACTTTCAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGAGCAGCAGGATTGG 240

DB 785 AAGAAGCACTTTCAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGAGCAGCAGGATTGG 844

QY 241 CTGAAGGAGAAAGTCAGAGCTGCAGAGCTGGCTGGCTGCCAACAACTCTCGTTGGCCAAA 300

DB 845 CTGAAGGAGAAAGTCAGAGCTGCAGAGCTGGCTGGCTGCCAACAACTCTCGTTGGCCAAA 904

QY 301 GCCAACAGCAGACACCTGGAGGATATGAACAGCCAGCTCAACTCATTTACAGTCCAGATG 360

DB 905 GCCAACAGCAGACACCTGGAGGATATGAACAGCCAGCTCAACTCATTTACAGTCCAGATG 964

QY 361 GAGAACATCACCACCTATCTCTCAAGCCACAGCAGCAAGACCTGAAAGACCTGCAGGACTTA 420
 DB 965 GAGAACATCACCACCTATCTCTCAAGCCACAGCAGCAAGACCTGAAAGACCTGCAGGACTTA 1024

QY 421 CACAAAGATGCAGAGATAGAACAGCCATCAAGTTCAACCAACTGGAGGACCTTCCAG 480
 DB 1025 CACAAAGATGCAGAGATAGAACAGCCATCAAGTTCAACCAACTGGAGGACCTTCCAG 1084

QY 481 CTCTTTGAGACGGATATTGTGAACATCATTAGCAATATCAGTTTACACAGCCACCACTG 540
 DB 1085 CTCTTTGAGACGGATATTGTGAACATCATTAGCAATATCAGTTTACACAGCCACCACTG 1144

QY 541 CGGACGCTGACCAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAA 600
 DB 1145 CGGACGCTGACCAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAA 1204

QY 601 CACACAGATGATGACCTTCCCTTGAATATACCTTGGCCAAACATCCGTTTGGATTCTGTT 660
 DB 1205 CACACAGATGATGACCTTCCCTTGAATATACCTTGGCCAAACATCCGTTTGGATTCTGTT 1264

QY 661 TCTCTCAGGATGCAACAAGATTGATGAGTGCAGGTTAGACACTCAAGTAGGCAACTTAA 720
 DB 1265 TCTCTCAGGATGCAACAAGATTGATGAGTGCAGGTTAGACACTCAAGTAGGCAACTTAA 1324

QY 721 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTTCCAAGCATGGTCAGCTCATCAAGAAT 780
 DB 1325 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTTCCAAGCATGGTCAGCTCATCAAGAAT 1384

QY 781 TTTACAAATACTACAAGTCCACCGGCCCCAGGGTCCCAAGAGGTACACAGATCCCCAG 840
 DB 1385 TTTACAAATACTACAAGTCCACCGGCCCCAGGGTCCCAAGAGGTACACAGATCCCCAG 1444

QY 841 GGACCCCTTGGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGGGGAGCGCTGGACCA 900
 DB 1445 GGACCCCTTGGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGGGGAGCGCTGGACCA 1504

QY 901 CTTGGCCCTTGGGGTGAGAGAGGCCCAATTTGGACCACTGGTCCCTCCCGGAGAGCGTGGC 960
 DB 1505 CTTGGCCCTTGGC----- 1516

QY 961 GGCAAAAGGATCTAAAGGCTCCAGGGCCCAAGAGGCTCCCGTGGTCCCTTGGGAAGCC 1020
 DB 1517 ----- 1516

QY 1021 GGCCCTCAGGGCCCTCAGTGGGACCCAGGCCCCCGGGCCACAGGCAAGAGGACTC 1080
 DB 1517 ----- 1516

QY 1081 CCGGGCCCTCAGGGCCCTCCTGGCTTCCAGGGACTTCAGGGACCCGTTGGGGAGCGTGGG 1140
 DB 1517 ----- 1516

QY 1141 GTGCCTTGGACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAGGCATGCCAGGCCCAAG 1200
 DB 1517 ----- 1516

QY 1201 GGCCCGCCCGCCCTCCTGGCCCATCAGGAGGGTGGTGCCTGGCCCTGCAGAGATGAG 1260
 DB 1517 ----- 1516

QY 1261 CCAACCCCGGACCGGAGGACAAATGGCTGCCCGCTCCTGGAAGAACTTCACAGACAAA 1320
 DB 1517 ----- 1552

QY 1321 TGCTACTATTTTTCAGTTTCAGAAAGAAATTTTGGAGATGCAAAAGCTTTCTGTGAAGAC 1380
 DB 1553 TGCTACTATTTTTCAGTTTCAGAAAGAAATTTTGGAGATGCAAAAGCTTTCTGTGAAGAC 1612

QY 1381 AAGTCTTTCACATCTCTTTTTCATTAACACTAGAGAGGAAACAGCAATGGATAAAAAACAG 1440
 DB 1613 AAGTCTTTCACATCTCTTTTTCATTAACACTAGAGAGGAAACAGCAATGGATAAAAAACAG 1672

QY 1441 ATGGTAGGGAGAGAGAGCCACTGGATCGGCCTCACAGACTCAGAGCGGTGAAAAATGAATGG 1500

|||||
Db 1673 ATGGTAGGAGAGAGCCACTGGATCGGCTCACAGACTCAGACGCTCAAAATGAATGG 1732
Qy 1501 AAGTGCTGGATGGGACATCTCCAGACTACAAAATTTGAAAGCTGGACGCGGATAC 1560
Db 1733 AAGTGCTGGATGGGACATCTCCAGACTACAAAATTTGAAAGCTGGACGCGGATAC 1792
Qy 1561 TGGGTCATGCGCATGGGCCAGGAGAGACTGTGCTGGTTGATTTATGCTGGGAGTGG 1620
Db 1793 TGGGTCATGCGCATGGGCCAGGAGAGACTGTGCTGGTTGATTTATGCTGGGAGTGG 1852
Qy 1621 AACGATTTCCATGTAAGAGCTCAATACTTCATTTCCGAAAAGAGAGGAGACAGTA 1680
Db 1853 AACGATTTCCATGTAAGAGCTCAATACTTCATTTCCGAAAAGAGAGGAGACAGTA 1912
Qy 1681 CTGTCATCTGCATTA 1695
Db 1913 CTGTCATCTGCATTA 1927
RESULT 13
AAS71133
ID AAS71133 standard; cdna: 1521 BP.
XX AC AAS71133;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #6937.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR P-PSDB; ABG06946.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX PS Claim 1; SEQ ID No 6937; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1521 BP; 386 A; 420 C; 425 G; 290 T; 0 other;
Query Match 52.1%; Score 883.8; DB 23; Length 1521;
Best Local Similarity 99.8%; Pred. No. 1.5e-214;
Matches 885; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 793 CAAGGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCAGGAGCCCTCGGC 852
Db 172 CAAGGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCAGGAGCCCTCGGC 231
Qy 853 CCACTGGCAACAGGGGACAGAAAGAGAGAGAGGGGAGCCTGGACCCTCGGC 912
Db 232 CCACTGGCAACAGGGGACAGAAAGAGAGAGAGGGGAGCCTGGACCCTCGGC 291
Qy 913 GGTGAGAGAGGCCCAATTGGACCACTGTCCCGCCGAGAGCGTGGCGCAAGAGATCT 972
Db 292 GGTGAGAGAGGCCCAATTGGACCACTGTCCCGCCGAGAGCGTGGCGCAAGAGATCT 351
Qy 973 AAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCCGGCCCTCAGGC 1032
Db 352 AAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCCGGCCCTCAGGC 411
Qy 1033 CCCAGTGGGGAGCCAGGCCGCCCGGGCCACACAGAGCAAGAGGAGACTCCCGGCCCTCAG 1092
Db 412 TCCAGTGGGGAGCCAGGCCGCCCGGGCCACACAGAGCAAGAGGAGACTCCCGGCCCTCAG 471
Qy 1093 GGCCTCTCTGCTTCCAGGGACTTCAGGGCACCGTTGGGAGCCCTGGGGTGGCTGGACCT 1153
Db 472 GGCCTCTCTGCTTCCAGGGACTTCAGGGCACCGTTGGGAGCCCTGGGGTGGCTGGACCT 531
Qy 1153 CGGGGACTGCCAGGCTTGCTGGGTACAGGCATGCCAGGCCCCAAGGGCCCCCGCCG 1212
Db 532 CGGGGACTGCCAGGCTTGCTGGGTACAGGCATGCCAGGCCCCAAGGGCCCCCGCCG 591
Qy 1213 CTTCTGGCCCATCAGGAGCGGTGTGCCCTTGGCCCTGTCAGAAATGAGCAACCCGGCA 1272
Db 592 CTTCTGGCCCATCAGGAGCGGTGTGCCCTTGGCCCTGTCAGAAATGAGCAACCCGGCA 651
Qy 1273 CCGGAGGACATGGCTGCCCGCTCAGTGGAGAACTTCACAGACAAATGCTACTATTTT 1332
Db 652 CCGGAGGACATGGCTGCCCGCTCAGTGGAGAACTTCACAGACAAATGCTACTATTTT 711
Qy 1333 TCAGTTGAGAAAGAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTCACAT 1392
Db 712 TCAGTTGAGAAAGAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTCACAT 771
Qy 1393 CTTGTTTTATTAACACTAGAGAGAACAGCAATGATGATAAAACAGATGTAGGGAGA 1452
Db 772 CTTGTTTTATTAACACTAGAGAGAACAGCAATGATGATAAAACAGATGTAGGGAGA 831
Qy 1453 GAGAGCCACTGGATCGGCTTCACAGACTCAGAGCGTGAATGAATGGAAGTGGCTGGAT 1512
Db 832 GAGAGCCACTGGATCGGCTTCACAGACTCAGAGCGTGAATGAATGGAAGTGGCTGGAT 891
Qy 1513 GGGACATCTCCAGACTACAAAATTTGGAAGCTGGACGCGGATGAATCTGGGTGATGGC 1572
Db 892 GGGACATCTCCAGACTACAAAATTTGGAAGCTGGACGCGGATGAATCTGGGTGATGGC 951
Qy 1573 CATGGCCAGGAGAGACTGTGCTGGTTGATTTATGCTGGCAGTGGAACTTTCCAA 1632
Db 952 CATGGCCAGGAGAGACTGTGCTGGTTGATTTATGCTGGCAGTGGAACTTTCCAA 1011
Qy 1633 TGTGAAGAGCTCAATACTTTCATTTGCGAAAAGACAGGGAGACAGT 1679
Db 1012 TGTGAAGAGCTCAATACTTTCATTTGCGAAAAGACAGGGAGACAGT 1058

| | |
|-----------|---|
| RESULT 14 | |
| AAAX27858 | |
| ID | AAAX27858 standard; DNA; 3685 BP. |
| XX | AAAX27858; |
| XX | |
| XX | 02-JUN-1999 (first entry) |
| XX | |
| XX | Human CSR3 protein coding sequence. |
| XX | |
| KW | Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage; |
| KW | scavenger receptor protein; intracellular stress; arteriosclerosis; |
| KW | diabetic circulatory obstruction; microbial infection; ss. |
| XX | |
| OS | Homo sapiens. |
| OS | |
| XX | WO9909159-A1. |
| XX | |
| PD | 25-FEB-1999. |
| XX | |
| XX | 12-AUG-1998; 98WO-JP03602. |
| XX | |
| PR | 30-JUL-1998; 98JP-0230121. |
| PR | 13-AUG-1997; 97JP-0233396. |
| XX | |
| PA | (NIBS) JAPAN TOBACCO INC. |
| XX | |
| PI | Nakamura Y, Tokino T; |
| XX | |
| DR | WPI: 1999-181032/15. |
| DR | P-PSDB; AA700994. |
| XX | |
| PT | Scavenger receptor proteins - for treatment and diagnosis of |
| PT | disorders involving cell stress |
| XX | |
| PS | Claim 6; Page 142-150; 175pp; Japanese. |
| XX | |
| CC | This sequence encodes the human cellular stress response 3 (CSR3) protein |
| CC | of the invention. The CSR proteins are macrophage scavenger receptor |
| CC | proteins. The CSR proteins can be used in the treatment, gene therapy |
| CC | and diagnosis of diseases in which intracellular stress is important, |
| CC | such as arteriosclerosis, diabetic circulatory obstruction, and microbial |
| CC | infection. Expression of the proteins is induced in vivo in response to |
| CC | intracellular stress, and inhibits cell death as a result of such stress. |
| XX | |
| SQ | Sequence 3685 BP; 901 A; 1142 C; 972 G; 669 T; 1 other; |
| | |
| | Query Match 14.6%; Score 248.2; DB 20; Length 3685; |
| | Best Local Similarity 50.5%; Pred. No. 1e-52; |
| | Matches 601; Conservative 0; Mismatches 588; Indels 0; Gaps 0 |
| OY | 42 TCTGCGAACCACCAATGTATTCTCATPAATGTGGTGCATCATGAACCTCAACACCTGAACCT 101 |
| | |
| Db | 765 TCTGAGGACCTCACCCAGGAGTGTACGATGTCAAGGCTGCATGTGCACCCAGATCAACTT 824 |
| OY | 102 GACCCAGGTGCAGCAGAGAACCTCATCAGAACTCTGCAGGGTCTCTGGATGACACAG 161 |
| | |
| Db | 825 CACCTGGGGGCGAGCTCCGAGTGGATCCACGGGATCCACGGGAAGACAGACGAGGAGAC 884 |
| OY | 162 CCAGGGTATCCAGCGAAATCAAGAAGCACTTTCAAAATCTGCAGCAGAGTTTTTCTCAAGC 221 |
| | |
| Db | 885 CTGACCTTCAGAGATTTGTCACCGACTGGCAGAACTACACAGGCTCTTCAGCGCCT 944 |
| OY | 222 CAAGAAGGACACGGATTGGCTGAAGGAGAAAGTGCAGAGCTTGCGACGCTGGCTGCCAA 281 |
| | |
| Db | 945 GCSCACCACTCCACCAAGACTTGGAGAGCGGTCAAGAACTATCCAGCCACCCCTGGGGG 1004 |
| OY | 282 CAACTCTGGTGGGCCAAGCCAAACAGCACCCTGGAGATATGAACAGCAGCTCAA 341 |
| | |
| Db | 1005 CTCTCAGACGGATCAGCCGAACTCAGAGAGATCAGCAGCTGGTACTCCAGGTCTAT 1064 |
| OY | 342 CTCATTTCAGAGTTCAGATGGAGAACTATCACCACCTATCTCTCAAGCCCAACGAGCAACCT 401 |

| Result No. | Score | Query | | Length | DB | ID | Description |
|------------|-------|-------|-------|--------|--------------------|--------------------|-------------|
| | | Match | % | | | | |
| 1 | 121.6 | 7.2 | 810 | 1 | US-08-642-255-60 | Sequence 60, Appl | |
| 2 | 118 | 7.0 | 756 | 1 | US-08-642-255-50 | Sequence 50, Appl | |
| 3 | 113.6 | 6.7 | 3181 | 1 | US-08-653-086-1 | Sequence 1, Appl | |
| c | 112.8 | 6.7 | 506 | 1 | US-08-253-155A-61 | Sequence 61, Appl | |
| 5 | 106 | 6.3 | 432 | 1 | US-08-642-255-48 | Sequence 48, Appl | |
| 6 | 106 | 6.3 | 1608 | 4 | US-09-029-348-19 | Sequence 19, Appl | |
| 7 | 105.4 | 6.2 | 503 | 1 | US-08-253-155A-20 | Sequence 20, Appl | |
| 8 | 100 | 5.9 | 5102 | 1 | US-08-494-168-1 | Sequence 1, Appl | |
| 9 | 99.6 | 5.9 | 1416 | 1 | US-07-621-091G-1 | Sequence 1, Appl | |
| 10 | 99.6 | 5.9 | 1416 | 2 | US-08-399-889-1 | Sequence 1, Appl | |
| 11 | 99.6 | 5.9 | 1416 | 3 | US-09-167-364-1 | Sequence 1, Appl | |
| 12 | 99.6 | 5.9 | 1416 | 4 | US-09-430-897-1 | Sequence 1, Appl | |
| 13 | 99.4 | 5.9 | 1868 | 1 | US-08-392-367B-1 | Sequence 1, Appl | |
| 14 | 99.4 | 5.9 | 1868 | 3 | US-08-893-467A-1 | Sequence 1, Appl | |
| 15 | 99.4 | 5.9 | 2543 | 1 | US-08-553-669-11 | Sequence 11, Appl | |
| 16 | 99.4 | 5.9 | 2543 | 3 | US-09-073-663-11 | Sequence 11, Appl | |
| 17 | 97.6 | 5.8 | 4359 | 4 | US-09-484-970B-4 | Sequence 4, Appl | |
| 18 | 96.2 | 5.7 | 1572 | 4 | US-09-297-269-39 | Sequence 39, Appl | |
| 19 | 95.2 | 5.6 | 1560 | 2 | US-08-794-795-5 | Sequence 5, Appl | |
| 20 | 95.2 | 5.6 | 1560 | 4 | US-09-249-200-5 | Sequence 5, Appl | |
| 21 | 95.2 | 5.6 | 1703 | 2 | US-08-794-795-1 | Sequence 1, Appl | |
| 22 | 95.2 | 5.6 | 1703 | 4 | US-09-249-200-1 | Sequence 1, Appl | |
| 23 | 89 | 5.3 | 1881 | 4 | US-09-029-348-20 | Sequence 20, Appl | |
| 24 | 77 | 4.5 | 1560 | 4 | US-09-453-702B-264 | Sequence 264, Appl | |
| 25 | 77 | 4.5 | 61663 | 4 | US-09-453-702B-62 | Sequence 62, Appl | |
| c | 76.8 | 4.5 | 5552 | 3 | US-08-155-888-1 | Sequence 1, Appl | |
| 27 | 74.6 | 4.4 | 38584 | 4 | US-09-453-702B-50 | Sequence 50, Appl | |


```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3181 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-655-086-1

Query Match          6.7%; Score 113.6; DB 1; Length 3181;
Best Local Similarity 52.5%; Pred. No. 1.6e-20;
Matches 248; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 766 CAGCTCATCAAGATTTTACAAATACAGGTCCACCGGGCCCGAGGGTCCAAAGAGGT 825
Db 104 TTGGTCCCAAGCTTCCCAAGGCTTCCAAAGGTCCTCCAGGCTGGAGCTGGAGCTGGA 163
QY 826 GACAGAGGATCCAGCGACCCCTCGCCCACTGGCCCAACAGGAGGACAGAAAGGAGAG 885
Db 164 GTTTCAGGTCCCATGGGTCCCGGAGGTCCCCCAGGTCCCCCTGGAAAGAAATGGAGAT 223
QY 886 GGGGAGCCTGGACCACTGGCCCTCGGGGTGAGAGAGGCCCAATTTGGACCACTGGTCCC 945
Db 224 GGGGAGCTGGAAACCTGGTCTCTGTGTAGCGTGGCCCTCTCTGGCCCTCAGGGTGT 283
QY 946 CCCGAGAGCGTGGCGGCAAGGATCTAAAGGCTCCCAAGGGCCCCCAAGGGTCCCGTGT 1005
Db 284 CGAGGATTCGCCGGAACAGCTGGCTCCCTGGAATGAAGGACACAGAGGTTTCAGTGT 343
QY 1006 TCCCTGGGAAGCCCGCCCTCAGGGCCCCAGTGGGGACCCAGGCCCGCCCGGCCCA 1065
Db 344 TTGGATGTGCAAGGGAGATGCTGGTCTCTGTGTCTTAAGGGTGAAGCTGGCAGCCCT 403
QY 1066 GGCAAGAGGGACTCCCGGCCCTCAGGGCCCTCTGGCTTCCAGGGACTTCAGGGCACC 1125
Db 404 GTGAAATAGGAGCTCTGGTTCAGATGGCCCCCGCTGGCTGGTGGAGAGAGTGGC 463
QY 1126 GTTGGGAGCGTGGGGTGGCTGGAGCTCGGGGACTCGCAGGGCTTCCCTGGGGTACCAG 1185
Db 464 CTTGAGGCCCTTGGCCCTGCTGGTCTGCTGGAAATGATGTGTCTACTGTCTGCCGG 523
QY 1186 ATGCCAGGCCCAAGGGCCCCCGCCCTCTCGGCCCATCAGGAGCGGTGG 1237
Db 524 CCCCTGGTCCACCGGCCCGCTGGTCTCTCTGCTCTCCCTGGTGTGGTGTG 575

RESULT 4
US-08-253-155A-61/c
; Sequence 61, Application US/08253155A
; Patent No. 5691147
; GENERAL INFORMATION:
; APPLICANT: Gyuris, Jeno
; APPLICANT: Draetta, Giulio
; TITLE OF INVENTION: CDK4 Binding Proteins
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/253.155A
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
```

```
;
; REFERENCE/DOCKET NUMBER: MII-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 506 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-253-155A-61

Query Match          6.7%; Score 112.8; DB 1; Length 506;
Best Local Similarity 53.8%; Pred. No. 1.1e-20;
Matches 254; Conservative 0; Mismatches 217; Indels 1; Gaps 1;

QY 506 TCATTAGCAATATCATGTTACACAGCCCACTGGGGACGTG-ACCAGCAATCTAAAT 564
Db 506 TCTTCAACAATCAATTTCCACCGACCACTGTCACAAACATGCTCCAAAGTACCTGGAT 447
QY 565 GAAGTCAGSACCACCTTGCACAGATACCTTTACCAACACACAGATGATCTGACCTCCTTG 624
Db 446 GACGTTCGGCTCTCTGACAGCTGGGCTTCCACACCCATGCGGAGAGCTCTACTACCTG 387
QY 625 AATAATACCTGGCCCAACATCCGTTTGGATTCTTCTCTCAGGATGCAACAAGATTTG 684
Db 386 RACAAGTGTGTCTCATGCTGGGCACACAGACCTGCTCCGGGAGCGCTTCAGCCTG 327
QY 685 ATGAGGTGAGGTTAGACACTGAAGTAGCAACTTATCATGATTATGGAAGAAATGAAG 744
Db 326 CTCATGTCGGCTGGACCTCAACGTCCGGAACCTCTCCATGATCGTGGAGAGATGAAG 267
QY 745 CTAGTAGACTCAACATGCTGAGCTCATCAAGAATTTTACAAATCTACAAGTCCACCG 804
Db 266 GCAGTGGACACACAGATGGAGAAATCTTCCGAATGTCCACCATCTACGAGGTGCCCC 207
QY 805 GCGCCAGGCTCCCAAGAGGTGACAGAGATCCCAAGGACCCCTGGCCCAACTGGCAAC 864
Db 206 GCGCTCCAGGACCAAGAGGATTCCAGAGATATGGGGCTCAAGAGCGCTTGTGGCGC 147
QY 865 AAGGACAGAAAGAGAGAGAGGGGAGCGCTGGACCACTGGCCCTGCGGGTGGAGAGGC 924
Db 146 AGAGCCCGGAAAGAGAGAGCCCGCAGCTTGGGCCCTCGGGACCCAGGGTCTCTCAGGGA 87
QY 925 CCAATTGGACAGCTGGTCCCGGAGAGCGTGGCGGCAAAAGGATCTAAAG 976
Db 86 CAACCTGGAGAGGCGCGGCTGTGGGAGAAAGGGGCCCTGTGTTGGCCCTCGAG 35

RESULT 5
US-08-642-255-48
; Sequence 48, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

US-09-029-348-19

Query Match 6.3%; Score 106; DB 4; Length 1608;
Best Local Similarity 52.8%; Pred. No. 1.3e-18;
Matches 229; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

Qy 796 GGTCACAGGGGCCCGAGGGGTCCAAGAGGTGACAGAGATCCAGGGACCCCCTGGCCCCA 855
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 GGTGAACCTGGTCAAACTGGTCCCTGCAGAGTGACCTGGTCTCATGGCCCGTGGGTCT 420

Qy 856 ACTGGCACAGGGGACAGAAAAGAGAGAAGGGGAGCCTGGACCACCTGGCCCTGGCGGT 915
||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 GCTGGCAAAACATGGAACCGGTGGTGAACACTGGTCTCTGGTCTCTGGTCTCTGGT 480

Qy 916 GAGAGAGCCCAAATTGGACCACTGGTCCCCCGGAGAGCGCTGGCGGCAAAAGGATCTAAA 975
| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 GCTGTTGGCCCAAGAGGTCTTAGTGGCCCAAGAGCATTCGTGGCGATAAGGAGAGCCC 540

Qy 976 GGCTCCAGAGGCCCGCAAGAGGTCCCGTGGTTCCTCGGTAAGCGCGGCCCTCAGGGCCCC 1035
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 GGTGAAGAGGGCCGACAGAGGTCTTCTGGCTCAAGGGACACAATGGATTGCAAGGTCTG 600

Qy 1036 AGTGGGACCCAGGCCCGCCGCGGCCACACAGGAAAGAGGACTCCCCGGCCCTCAGGGC 1095
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 CTTGGTATCGTGGTGACATGGTGTCAAGGTGCTCTCGTGGTCTCTGGTCTCTGGT 660

Qy 1096 CCTCGTGGCTTCCAGGGACTTCAGGGCACCTTTGGGGAGCGCTGGGGTGCCTGGACCTCGG 1155
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 CTAGGGGCCCTGCTGGTCTCTGGCCCTGCTGSAAGATGGTTCGACATGGACATCCT 720

Qy 1156 GGACTGCGAGCTTGGCTGGGGTACCAGGCATGCCAGGCCCGCCCGCCCGGCCCT 1215
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 GGTACGGTTGACCTGCTGGCATTCGAGGGCCCTCAGGGTCAACCAAGGCCCTCTGGCCCC 780

Qy 1216 CCTGCCCATCAGG 1229
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 781 CTGGTCCCCCTGG 794

RESULT 7
US-08-253-155A-20
Sequence 20, Application US/08253155A
Patent No. 5691147
GENERAL INFORMATION:
APPLICANT: Gyuris, Jenö
APPLICANT: Draetta, Giulio
TITLE OF INVENTION: CDK4 Binding Proteins
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/253.155A
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MII-028
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:

```
; ; LENGTH: 503 base pairs
; ; TYPE: nucleic acid
; ; STRANDEDNESS: single
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: cdna
US-08-253-155A-20

Query Match
Best Local Similarity 6.2%; Score 105.4; DB 1; Length 503;
Matches 226; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 387 CAACGAGCAGACCTGAAGACCTGCAGACCTTACCAAGATGCGAGAAATAGAACACG 446
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15 CCACGAGAGACATGATGCTTTCAGTACCATACCCACTACGCCAGAACCGCACTGT 74

QY 447 CATCAAGTTCAACCAACTGAGAGAGCTTCCAGCTCTTTGAGACGATATTGTGAACAT 506
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 GGAGAGTTTGAGTCTCTGGTAGGAGCATGGTCTTCACAGAGATTGAAATTGGCAACAT 134

QY 507 CATTAGCAATATCAGTTACACAGCCACCACCTCGCGAGCGTGACCAATCTTAATGA 566
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135 CTTCAACCAATCAATGCCACCGACACCAACACCGCGCACACATGCTCATGTACCTGGATGA 194

QY 567 AGTCAGGACCACTTGACAGATACCTTTACCAACACACAGATGATCTGACCTCTCTTGA 626
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 195 CGTGGCGCTCTCTGACACCTGGGCTTCACACACCCATGCCGAGGAGCTCTACTACCTGAA 254

QY 627 TAATACCTTGGCGCAACATCGTTTGGATCTGTCTCTCAGATGCAACAAGATTGTGAT 686
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 255 CRAAGTCTGTCTCATCATGCTGGGCACACAGACCTGCTCCGGAGCGGCTTCAGCCTGTCT 314

QY 687 GAGGTCGAGTTAGACACTGAAGTAGCCAACTTATCAGTGTGATTATGGAAGAAATGAAGCT 746
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 315 CAGTGGCGGCTGGACCTCAAGCTCGGGAACCTCTCCATGATCGTGGAGGATGAAGGG 374

QY 747 AGTAGACTCCAGCAGCTGCTAGCTCATCAAGAAATTTTACAATACTACAAGGTCCACCGGG 806
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 375 AGGGGACACACAGAATGGGGAGATCCTTCGGAATGTAAATCTTACAGATCTCTAGAGGTGCCCGCGC 434

QY 807 CCCCAGG 813
    |||||
Db 435 CTCACGG 441

RESULT 8
US-08-494-168-1
; Sequence 1, Application US/08494168
; Patent No. 5731192
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T.
; APPLICANT: Zhou, Jing
; TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method
; TITLE OF INVENTION: of Detecting Collagen Deficiency
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/494,168
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,465
; FILING DATE: 27-AUG-1993
; ATTORNEY/AGENT INFORMATION:
```

```
; ; NAME: SAXE, Bernhard D.
; ; REGISTRATION NUMBER: 28,665
; ; REFERENCE/DOCKET NUMBER: 40397/104/BABR
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: (202)672-5300
; ; TELEFAX: (202)672-5399
; ; TELEX: 904136
; ; INFORMATION FOR SEQ ID NO: 1:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 5102 base pairs
; ; TYPE: nucleic acid
; ; STRANDEDNESS: double
; ; TOPOLOGY: linear
; ; FEATURE:
; ; NAME/KEY: CDS
; ; LOCATION: join(2..82, 86..97, 101..4399, 4403..4420, 4424
; ; LOCATION: ..4465, 4469..4876, 4880..5101)
US-08-494-168-1

Query Match
Best Local Similarity 5.9%; Score 100; DB 1; Length 5102;
Matches 220; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 792 ACAAGTCCACCGGGCCCCCAGGGTCCAAAGAGGTGACAGAGGTACAGAGGATCCAGGGACCCCTGG 851
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2380 AAATGTTGCTCCGGGGGAAACAAGGCTTACAAGGATTAACAGGGGCACAAAGGATTTCTTGG 2439

QY 852 CCCAAGTCCCAACAGGGGACAGAAAGAGAGAGAGGGGAGCGCTGGACACCTGGCCCTGC 911
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2440 AGACTTGGCGCTTCCAGGACTCAAGGTTGTCACGGGAGGCTTACTTAGGCCCA 2499

QY 912 GGGTGAGAGAGGCCCAATTTGGACACAGCTGTCCTCCCGGAGAGCGTGGCGCAAAAGATC 971
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2500 AGGTGAGCGGGGACGCCCTTGGGACACCCAGGACAGGTGGACAGCCAGCCACCCAGATC 2559

QY 972 TAAAGCTCCAGGGGGCCCCAAAGGCTCCCGTGGTTTCCCTTGGGAAGCCCGGCCCTCAGG 1031
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2560 TAGTGTCTCATATGTCATCAAGGGCAATCTGGGCTTCCAGGAGCACCAGGCTTCCAGG 2619

QY 1032 CCCCAGTGGGACCCAGGCCCCCGGCCCCCAGGCAAGAGGGGACTCCCGGCCCTCA 1091
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2620 CATCTCAGGACATCTCTGGAAGAAAGAACAGAGGCAAGAAAGGTCTCTCTGGATCAAT 2679

QY 1092 GGGCCCTCTGCTTCCAGGGACTTCAGGGCACCTTGGGGAGCCCTGGGTGCTGGACC 1151
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2680 TGTAAAGAAAGGCTGCCAGGCTAAAGGCTTCTCTGGAATCCAGGCTTAGTAGACT 2739

QY 1152 TCGGGGACTGCAGGCTTGCCTGGGTACCAAGCATGCCAGGCATGCCAGGCCCAAGGCCCGCCG 1211
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2740 GAAAGAGCCAGGCTCTCCAGGGTCTCGTGGTTCGCCAGGCCCTCTCTGGACCAAGG 2799

RESULT 9
US-07-621-091G-1
; Sequence 1, Application US/07621091G
; Patent No. 542408
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T., Morrison, Karen E., Hudson, Billy
; APPLICANT: G.
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen
; TITLE OF INVENTION: Polynucleotides
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University, Office of Cooperative Research
; STREET: 246 Church Street
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800K storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS7.0
```

SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/621,091G
FILING DATE: 11/30/90
CLASSIFICATION: 424
PRIOR APPLICATION DATA: No. 5424408 applicable
ATTORNEY/AGENT INFORMATION:
NAME: Barth, Richard S.
REGISTRATION NUMBER: 28180
REFERENCE/DOCKET NUMBER: 900983/RB
TELEPHONE: (212) 972-1400
TELEPHONE: (212) 972-1400
TELEPHONE: (212) 370-1622
TELEPHONE: (212) 370-1622

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1416 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Calf
STRAIN: Unknown
INDIVIDUAL ISOLATE: Unknown
DEVELOPMENTAL STAGE: Unknown
CELL TYPE: Whole kidney
IMMEDIATE SOURCE:
LIBRARY: Bovine lens cDNA
CLONE: KMC15
POSITION IN GENOME: No. 5424408 known
CHROMOSOME/SEGMENT: No. 5424408 known
PUBLICATION INFORMATION: No. 5424408e
US-07-621-091G-1

Query Match 5.9%; Score 99.6; DB 1; Length 1416;
Best Local Similarity 51.6%; Pred. No. 5.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 796 GGTCCACCGGGCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCAGGAGCCCCCTGCCCCA 855
Db 283 GGAGCACCAGGTGTCCCGGTGACAGGGGGCCAGAGGTGATCCTGGATTCTATGATT 342

QY 856 ACTGGCAACAAGGACAG 915
Db 343 CGAGGATGAAG 402

QY 916 GAGAGAGGCCCCAATTGGACAGCTGGTCCCGCCGAGAGAGCTGGCGGCAAAAGGATCTAA 975
Db 403 CAAAGTGGACCAAAAGGACCACTGGAGTACGTGGAGAGCCTGGCACAGTGAAGATCATC 462

QY 976 GGTCCACCGGGCCCCAAGGCTCCCTGGTTCCTCCCTGGAGAGCCCGGCCCTCAGGCCCC 1035
Db 463 TCCTCTCCAGAGAGCCAGGCCACCTGCTGAGTTCAGGATTTCCAGGACCACTGGACCTCAGG 522

QY 1036 AGTGGGACCCAGGCCCCCGGGCCCAACAGGCAAAAGAGAGAGTCCCGGCCCTCAGGCG 1095
Db 523 CCGGGCCCCAGGACCACTGGAGTACGTGGAGAGCCTGGCACAGTGAAGATCATC 582

QY 1096 CCTCTGGCTTCAGGAGCTTCAGGACCTTCAGGACCTGGGGAGCTGGGCTGCTGACCTCG 1155
Db 583 GAGGATGGTCCACCAAGCACTCTGGACCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 642

QY 1156 GGACTGCCAGGCTTGCTGGGTGATCCAGGATCCAGGAGCCCAAGGCCCCCCCCCGCCCT 1215
Db 643 GGAGAGCAAGGACCACTGGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCG 702

QY 1216 CTGGGCCCATCAGGAGCGGTGG 1237
Db 703 ACTGGACCACTGCACGAGGGG 724

RESULT 11

US-09-167-364-1
; Sequence 1, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeder, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B

RESULT 10
US-08-399-889-1
; Sequence 1, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeder, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Calf
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
US-08-399-889-1

Query Match 5.9%; Score 99.6; DB 2; Length 1416;
Best Local Similarity 51.6%; Pred. No. 5.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 796 GGTCCACCGGGCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCAGGAGCCCCCTGCCCCA 855
Db 283 GGAGCACCAGGTGTCCCGGTGACAGGGGGCCAGAGGTGATCCTGGATTCTATGATT 342

QY 856 ACTGGCAACAAGGACAG 915
Db 343 CGAGGATGAAG 402

QY 916 GAGAGAGGCCCCAATTGGACAGCTGGTCCCGCCGAGAGAGCTGGCGGCAAAAGGATCTAA 975
Db 403 CAAAGTGGACCAAAAGGACCACTGGAGTACGTGGAGAGCCTGGCACAGTGAAGATCATC 462

QY 976 GGTCCACCGGGCCCCAAGGCTCCCTGGTTCCTCCCTGGAGAGCCCGGCCCTCAGGCCCC 1035
Db 463 TCCTCTCCAGAGAGCCAGGCCACCTGCTGAGTTCAGGATTTCCAGGACCACTGGACCTCAGG 522

QY 1036 AGTGGGACCCAGGCCCCCGGGCCCAACAGGCAAAAGAGAGAGTCCCGGCCCTCAGGCG 1095
Db 523 CCGGGCCCCAGGACCACTGGAGTACGTGGAGAGCCTGGCACAGTGAAGATCATC 582

QY 1096 CCTCTGGCTTCAGGAGCTTCAGGACCTTCAGGACCTGGGGAGCTGGGCTGCTGACCTCG 1155
Db 583 GAGGATGGTCCACCAAGCACTCTGGACCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 642

QY 1156 GGACTGCCAGGCTTGCTGGGTGATCCAGGATCCAGGAGCCCAAGGCCCCCCCCCGCCCT 1215
Db 643 GGAGAGCAAGGACCACTGGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCG 702

QY 1216 CTGGGCCCATCAGGAGCGGTGG 1237
Db 703 ACTGGACCACTGCACGAGGGG 724

;; CURRENT APPLICATION NUMBER: US/09/167.364
;; CURRENT FILING DATE: 1998-10-07
;; EARLIER APPLICATION NUMBER: 08/399889
;; EARLIER FILING DATE: 1995-03-07
;; NUMBER OF SEQ ID NOS: 25
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 1416
;; TYPE: DNA
;; ORGANISM: Calf
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(1416)
US-09-167-364-1

Query Match 5.9%; Score 99.6; DB 3; Length 1416;
Best Local Similarity 51.6%; Pred. No. 5.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 796 GGTCCACCGGGCCCCCAGAGGTCCAAGAGGTGACAGAGTCCAGGAGCCCTCGGCCCA 855
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 283 GGAGCACCAGGTGTCCCGGTGACGAGGCGCCAGAGGTGATCCTGGATTCTATGATT 342
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 856 ACTGCGACACAGGACACAGAAAGAGAGAGAGGAGGAGCTGGACCACTGGCCCTCGGGT 915
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 343 CCAGGCATGAAGGAAGAGAGGTAATTCAGGATTTCCAGGACCACTGGACCTCCAGG 402
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 916 GAGAGAGGCCCAATTGGACCACTGGTCCCGCCGAGAGCGTGGCGGCAAGGATCTAAA 975
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 403 CAAAGTGACCAAAAGGACCACTGGATTCAGGATTTCCAGGACCACTGGACCTCCAGG 462
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 976 GGTCTCCAGGGCCCCCAAGGCTCCGTTCCCTGGGAAGCCGCGCCCTCAGGGCCCC 1035
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 463 TCCCTTCCAGGAAGCCAGGCCCACTGGTTCCAGTGGAGAACAGGATGAAGATCATC 522
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 1036 AGTGGGACCCAGGCCCCCGGGCCCAAGAGGAGTCCCGGAGGAGTCCCGGCGCCCTCAGGGC 1095
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 523 CCGGGCCCCAGGACCCAGGCCCACTGGTTCCAGTGGAGAACAGGATGAAGAGAA 582
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 1096 CCTCTCGGCTTCCAGGCACTTCAGGCGACCGTGGGAGCCCTGGGCTGCTGACCTCG 1155
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 583 GAGGATGTCTCCAGGAACTCTGAGCACTGGAGAACAGGATGAAGATGAAGATGA 642
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 1156 GGACTGCCAGGCTTGCCTGGGTACCAAGCATGCCAGGCCCCCAAGGGCCCCCGGCCCT 1215
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 643 GGAGAGCAAGGACCACTGGATCCGATGCCCTGCCAGGCTTGAAGGGGAAACCTTGAGAC 702
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 1216 CTGCGCCCATCAGGAGCGGTGG 1237
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 703 ACTGGACCACTGCAGCAGGGG 724
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 12
US-09-439-897-1
;; Sequence 1, Application US/09439897
;; Patent No. 6277558
;; GENERAL INFORMATION:
;; APPLICANT: Hudson, Billy G
;; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
;; FILE REFERENCE: 95-1263-C
;; CURRENT APPLICATION NUMBER: US/09/439.897
;; CURRENT FILING DATE: 1999-11-12
;; NUMBER OF SEQ ID NOS: 65
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 1416
;; TYPE: DNA
;; ORGANISM: Bos taurus
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(1416)
US-09-439-897-1

Query Match 5.9%; Score 99.6; DB 4; Length 1416;
Best Local Similarity 51.6%; Pred. No. 5.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 796 GGTCCACCGGGCCCCCAGAGGTCCAAGAGGTGACAGAGTCCAGGAGCCCTCGGCCCA 855
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 283 GGAGCACCAGGTGTCCCGGTGACGAGGCGCCAGAGGTGATCCTGGATTCTATGATT 342
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 856 ACTGCGACACAGGACACAGAAAGAGAGAGAGGAGCCTGGACCACTGGCCCTCGGGCT 915
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 343 CCAGGCATGAAGGAAGAGGTAATTCAGGATTTCCAGGACCACTGGACCTCCAGG 402
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 916 GAGAGAGGCCCAATTGGACCACTGGTCCCGCCGAGAGCGTGGCGGCAAGGATCTAAA 975
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 403 CAAAGTGACCAAAAGGACCACTGGATTCAGGATTTCCAGGACCACTGGACCTCCAGG 462
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 976 GGTCTCCAGGGCCCCCAAGGCTCCGTTCCCTGGGAAGCCGCGCCCTCAGGGCCCC 1035
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 463 TCCCTTCCAGGAAGCCAGGCCCACTGGTTCCAGTGGAGAACAGGATGAAGAGAA 522
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 1036 AGTGGGACCCAGGCCCCCGGGCCCAAGAGGAGTCCCGGAGGAGTCCCGGCGCCCTCAGGGC 1095
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 523 CCGGGCCCCAGGACCCAGGCCCACTGGTTCCAGTGGAGAACAGGATGAAGAGAA 582
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 1096 CCTCTCGGCTTCCAGGCACTTCAGGCGACCGTGGGAGCCCTGGGCTGCTGACCTCG 1155
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 583 GAGGATGTCTCCAGGAACTCTGAGCACTGGAGAACAGGATGAAGATGAAGATGA 642
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 1156 GGACTGCCAGGCTTGCCTGGGTACCAAGCATGCCAGGCCCCCAAGGGCCCCCGGCCCT 1215
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 643 GGAGAGCAAGGACCACTGGATCCGATGCCCTGCCAGGCTTGAAGGGGAAACCTTGAGAC 702
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 1216 CTGCGCCCATCAGGAGCGGTGG 1237
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 703 ACTGGACCACTGCAGCAGGGG 724
|| |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 13
US-08-392-367B-1
;; Sequence 1, Application US/08392367B
;; Patent No. 5691197
;; GENERAL INFORMATION:
;; APPLICANT: Tryggvason, Karl
;; APPLICANT: Elomaa, Outi
;; APPLICANT: Kangas, Maarit
;; TITLE OF INVENTION: An Insolated DNA Sequence For a
;; Patent No. 5691197
;; TITLE OF INVENTION: No. 5691197el Macrophage Receptor with
;; TITLE OF INVENTION: a Collagenous Domain and the
;; TITLE OF INVENTION: Polypeptide Chain Encoded by
;; TITLE OF INVENTION: such a Sequence
;; NUMBER OF SEQUENCES: 2
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
;; ADDRESSEE: Minnich & McKee
;; STREET: 1100 Superior Avenue
;; STREET: Suite 700
;; CITY: Cleveland
;; STATE: Ohio
;; COUNTRY: U.S.A.
;; ZIP: 44114-2518
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 inch,
;; MEDIUM TYPE: 720 Kb storable
;; COMPUTER: IBM PS/2, Model 35 SX
;; OPERATING SYSTEM: DOS 5.0
;; SOFTWARE: Word Perfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/392.367B
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Minnich, Richard J.

Sequence 11, Application US/08555669
Patent No. 5773248
GENERAL INFORMATION:
APPLICANT: Brewton, Richard G.
APPLICANT: Mayne, Richard
TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/555,669
FILING DATE: 13-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 8389-030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-3660
TELEFAX: 415-854-3694
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2543 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 47..2098
US-08-555-669-11

Query Match 5.9%; Score 99.4; DB 1; Length 2543;
Best Local Similarity 51.5%; Pred. No. 8.e-17;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

| | | | |
|----|------|--|------|
| QY | 793 | CAAGTCCACCGGGCCCCAGGGTCCAGAGGTGACAGAGATCCAGGGACCCCTGGC | 852 |
| DB | 584 | CAAGTCCACCGGGCCCCAGGGTCCAGAGGTGACAGAGATCCAGGGACCCCTGGC | 643 |
| QY | 853 | CCAAGTCCACCGGGCCCCAGGGTCCAGAGGTGACAGAGATCCAGGGACCCCTGGC | 912 |
| DB | 644 | GAGCAGGGGGAGTCCGCAAGGACGGCGAGAGAGGTGACCTGGGCCCCCTGGGCCCC | 703 |
| QY | 913 | GGTGAGAGAGGCCCAATTGGACAGCTGCTCCCCCGGAGAGCGTGGCGCAAGAGATCT | 972 |
| DB | 704 | GGCTCCCGGCGAGCGTGGGCTGACGGCCCCCGGGGATTACAGAGGATGCCAGGGCCA | 763 |
| QY | 973 | AAAGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCTGGGAGCCCGGCCCTCAGGGC | 1032 |
| DB | 764 | CTCGGGCCCCCTGGGGACCGGGTCCCATGGGTTCCGAGGGCCGCTGGGATCCAGGA | 823 |
| QY | 1033 | CCAGTGGGGACCCAGGGCCCCCGGGCCACAGGCAAGAGGACTCCCGGGCCCTCAG | 1092 |
| DB | 824 | GGCTGGGAAGCGGGTGACCGAGGAGAGAGGGGCCAGAGAGGTTCGCGGGCCCCAAG | 883 |
| QY | 1093 | GGCCCTCCTGGGCTCCAGGGACTTCAGGGCACCGTTGGGGAGCTTGGGTGCTGGACCT | 1152 |
| DB | 884 | GGTGACCTCGGAGACCTGGTCCCAAGGGAACCCCGGAGTGGCGGGCCAGCGGAGAG | 943 |
| QY | 1153 | CGGGGACTGCGAGGCTTCCTGGGGTACCAAGCATGCCAGGCCCCCAAGGGCCCCCGGC | 1212 |

Db 944 CGGGCATGCCAGGCAAGGACGGCCAGAAATGCGTGCCAGGACTCGATGGCCAGAGGGA 1003
QY 1213 CTTCTGTGGCCATCAGAGCGGTGG 1237
Db 1004 GAGGCTGGTCCCAACGCGTGTCTCCG 1028
Search completed: March 21, 2003, 08:25:28
Job time : 79.7466 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 00:58:31 ; Search time 269.408 Seconds
(without alignments)
11326.533 Million cell updates/sec

Title: US-09-763-712a-1_COPY_670_2024

Perfect score: 1355

Sequence: 1 atgcaacaagattgatgag.....agaaaagggttatcatcccg 1355

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | ID | Description |
|------------|--------|-------|-------|--------|----------|--------------------|
| 1 | 1355 | 100.0 | 2024 | 21 | AAA07697 | Human collectin en |
| 2 | 1355 | 100.0 | 2628 | 22 | AAH43036 | Nucleotide sequenc |
| 3 | 1337.4 | 98.7 | 2005 | 24 | ABL95574 | Human anglogenesis |
| 4 | 1337.4 | 98.7 | 2053 | 24 | ABL88085 | Human PRO7223 cDNA |
| 5 | 1337.4 | 98.7 | 2318 | 22 | AA158842 | Human polynucleoti |
| 6 | 1337.4 | 98.7 | 2641 | 22 | AA158842 | Human EXMAD-14 cod |
| 7 | 1335.8 | 98.6 | 2930 | 20 | AAV55746 | Human secreted pro |
| 8 | 1328.8 | 98.1 | 2181 | 22 | AA160628 | Human polynucleoti |
| 9 | 1323.8 | 97.7 | 2929 | 24 | ABQ92072 | Human polynucleoti |

| | | | | | | |
|----|-------|------|------|----|----------|--------------------|
| 10 | 1029 | 75.9 | 2262 | 24 | ABA97932 | Human scavenger re |
| 11 | 895.2 | 66.1 | 2637 | 22 | AAH43037 | Nucleotide sequenc |
| 12 | 883.8 | 65.2 | 1521 | 23 | AAH71133 | DNA encoding novel |
| 13 | 743.8 | 54.9 | 2256 | 22 | AAH43054 | Nucleotide sequenc |
| 14 | 413 | 30.5 | 493 | 23 | AAH71131 | DNA encoding novel |
| 15 | 332 | 24.5 | 5041 | 22 | ABAL7153 | Human nervous syst |
| 16 | 249 | 18.4 | 415 | 24 | ABL66665 | Lung cancer relate |
| 17 | 157.6 | 11.6 | 3685 | 20 | AAH27858 | Human CSR1 protein |
| 18 | 157.6 | 11.6 | 3810 | 20 | AAH27856 | Human CSR1 protein |
| 19 | 150.2 | 11.1 | 873 | 23 | AAH71130 | DNA encoding novel |
| 20 | 150.2 | 11.1 | 1062 | 23 | AAH67442 | DNA encoding novel |
| 21 | 146.8 | 10.8 | 1329 | 23 | AAH71134 | DNA encoding novel |
| 22 | 121.6 | 9.0 | 756 | 14 | AAQ43034 | Collagen-like poly |
| 23 | 121.6 | 9.0 | 756 | 17 | AAH16768 | Collagen-like poly |
| 24 | 121.4 | 9.0 | 5676 | 15 | AAO64556 | Human collagen (TY |
| 25 | 121.4 | 9.0 | 8284 | 22 | AAH26526 | Human breast cance |
| 26 | 121.4 | 9.0 | 8284 | 22 | AAH26553 | Human breast cance |
| 27 | 121.4 | 9.0 | 8284 | 22 | AAH26600 | Human breast cance |
| 28 | 118.4 | 8.7 | 6109 | 23 | AAH78667 | Murine Col3a3 cDNA |
| 29 | 118 | 8.7 | 756 | 14 | AAQ43032 | Collagen-like poly |
| 30 | 117 | 8.6 | 756 | 17 | AAH16766 | Collagen-like poly |
| 31 | 117 | 8.6 | 9287 | 24 | ABK64501 | Human benign prost |
| 32 | 115.8 | 8.5 | 4748 | 22 | AAO66573 | Bovine alpha1(I) c |
| 33 | 115 | 8.5 | 4270 | 24 | AB199885 | Mouse ischaemic co |
| 34 | 114.8 | 8.5 | 4816 | 24 | ABK84041 | Human cDNA differe |
| 35 | 114.8 | 8.5 | 5058 | 23 | AAH69026 | DNA encoding novel |
| 36 | 114.8 | 8.5 | 6358 | 22 | AAH98286 | Human EST-derived |
| 37 | 114.8 | 8.5 | 6691 | 23 | AAH79806 | DNA encoding novel |
| 38 | 114 | 8.4 | 5060 | 24 | AAH92176 | Prostate cancer-as |
| 39 | 113.6 | 8.4 | 3181 | 19 | AAV59338 | Nucleotide sequenc |
| 40 | 113.4 | 8.4 | 4821 | 23 | AAH86866 | DNA encoding novel |
| 41 | 113.4 | 8.4 | 6200 | 23 | AAH78668 | Human COL5A3 cDNA |
| 42 | 113.4 | 8.4 | 6368 | 24 | ABN59647 | Novel human coding |
| 43 | 113.2 | 8.4 | 2319 | 13 | AAO24551 | DNA encoding human |
| 44 | 113.2 | 8.4 | 2319 | 18 | AAH47812 | Human type IV coll |
| 45 | 113.2 | 8.4 | 3898 | 22 | AAH76377 | Murine adipocytes- |

ALIGNMENTS

| | |
|----------|--|
| RESULT 1 | |
| AAA07697 | ID AAA07697 standard; DNA; 2024 BP. |
| XX | AC AAA07697; |
| XX | DF 20-JUN-2000 (first entry) |
| XX | DE Human collectin encoding DNA. |
| XX | DE Collectin; human; antibacterial; antiviral; ds. |
| KW | OS Homo sapiens. |
| XX | XX Key Location/Qualifiers |
| FT | CDS 55..1698 |
| FT | FT /*tag= a |
| FT | FT /product= "collectin" |
| FT | FT misc_feature 1..738 |
| FT | FT /*tag= b |
| FT | FT /note= "fragment specifically claimed in claim 8" |
| FT | FT misc_feature 55..738 |
| FT | FT /*tag= c |
| FT | FT /note= "fragment specifically claimed in claim 8" |
| FT | FT misc_feature 79..738 |
| FT | FT /*tag= d |
| FT | FT /note= "fragment specifically claimed in claim 8" |
| FT | FT misc_feature 325..738 |
| FT | FT /*tag= e |
| FT | FT /note= "fragment specifically claimed in claim 8" |
| FT | FT misc_feature 358..738 |
| FT | FT /*tag= f |

| | | | |
|----------------------------|---|---|------|
| Db | 2594 | TGGAACACTCCAATCAGAAAAAGTTATCATCCCG | 2628 |
| RESULT 3 | | | |
| ABL95574 | | | |
| ID | ABL95574 | standard; cDNA; 2005 BP. | |
| XX | XX | | |
| AC | ABL95574; | | |
| XX | XX | | |
| DT | 19-JUL-2002 | (first entry) | |
| XX | XX | | |
| DE | Human angiogenesis related cDNA PRO7223 | SEQ ID NO: 27. | |
| XX | XX | | |
| KW | Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer; | | |
| KW | atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder; | | |
| KW | cardiant; cytostatic; antiangiogenic; hypotensive; vulnerary; | | |
| KW | antiarteriosclerotic; gene; ss. | | |
| XX | XX | | |
| OS | Homo sapiens. | | |
| XX | XX | | |
| PN | WO200208284-A2. | | |
| XX | XX | | |
| PD | 31-JAN-2002. | | |
| XX | XX | | |
| PF | 09-JUL-2001; 2001WO-US21735. | | |
| XX | XX | | |
| PR | 20-JUL-2000; 2000US-219556P. | | |
| PR | 25-JUL-2000; 2000US-220624P. | | |
| PR | 25-JUL-2000; 2000US-220664P. | | |
| PR | 28-JUL-2000; 2000WO-US20710. | | |
| PR | 02-AUG-2000; 2000US-222695P. | | |
| PR | 17-AUG-2000; 2000US-0643657. | | |
| PR | 23-AUG-2000; 2000WO-US23322. | | |
| PR | 24-AUG-2000; 2000WO-US23328. | | |
| PR | 07-SEP-2000; 2000US-230978P. | | |
| PR | 15-SEP-2000; 2000US-000000P. | | |
| PR | 18-SEP-2000; 2000US-0664610. | | |
| PR | 18-SEP-2000; 2000US-0665350. | | |
| PR | 24-OCT-2000; 2000US-242922P. | | |
| PR | 08-NOV-2000; 2000US-0709238. | | |
| PR | 08-NOV-2000; 2000WO-US30952. | | |
| PR | 10-NOV-2000; 2000WO-US30873. | | |
| PR | 01-DEC-2000; 2000US-0522678. | | |
| PR | 20-DEC-2000; 2000US-0747259. | | |
| PR | 22-DEC-2000; 2000WO-US34956. | | |
| PR | 22-JAN-2001; 2001US-0767609. | | |
| PR | 28-FEB-2001; 2001US-0796498. | | |
| PR | 28-FEB-2001; 2001WO-US06520. | | |
| PR | 01-MAR-2001; 2001WO-US06666. | | |
| PR | 09-MAR-2001; 2001US-0802706. | | |
| PR | 14-MAR-2001; 2001US-0808689. | | |
| PR | 22-MAR-2001; 2001US-0816744. | | |
| PR | 05-APR-2001; 2001US-0828366. | | |
| PR | 10-MAY-2001; 2001US-0854208. | | |
| PR | 20-MAY-2001; 2001US-0854280. | | |
| PR | 25-MAY-2001; 2001US-0866028. | | |
| PR | 25-MAY-2001; 2001US-0866034. | | |
| PR | 25-MAY-2001; 2001WO-US17092. | | |
| PR | 30-MAY-2001; 2001US-0870574. | | |
| PR | 30-MAY-2001; 2001WO-US17443. | | |
| PR | 01-JUN-2001; 2001WO-US17800. | | |
| PR | 20-JUN-2001; 2001WO-US19692. | | |
| PR | 28-JUN-2001; 2001WO-US00000. | | |
| XX | XX | | |
| PA | (GETH) GENENTECH INC. | | |
| PA | (BAKE/) BAKER K P. | | |
| PA | (FERR/) FERRARA N. | | |
| PA | (GERB/) GERBER H. | | |
| PA | (GERR/) GERRITSEN M E. | | |
| PA | (GODD/) GODDARD A. | | |
| PA | (GODO/) GODOWSKI P J. | | |
| PA | (GURN/) GURNEY A L. | | |
| PA | (HILL/) HILLAN K J. | | |
| PA | (MARS/) MARSTERS S A. | | |
| PA | (PANJ/) PAN J. | | |
| PA | (PAON/) PAONI N F. | | |
| PA | (STEP/) STEPHAN J F. | | |
| PA | (WATA/) WATANABE C K. | | |
| PA | (WILL/) WILLIAMS P M. | | |
| PA | (WOOD/) WOOD W I. | | |
| XX | XX | | |
| PI | Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A; | | |
| PI | Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF; | | |
| PI | Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W; | | |
| XX | XX | | |
| DR | WPI; 2002-171999/22. | | |
| XX | XX | | |
| DR | P-PSDB; ABB95436. | | |
| XX | XX | | |
| PT | One hundred and eighty seven nucleic acids encoding PRO polypeptides, | | |
| PT | useful in diagnosis and treatment of cardiovascular (e.g. myocardial | | |
| PT | infarction), endothelial or angiogenic disorders in a mammal - | | |
| XX | XX | | |
| PS | Claim 1; Fig 27; 567pp; English. | | |
| XX | XX | | |
| CC | The present invention provides the protein and coding sequences of human | | |
| CC | PRO proteins. These are useful for treating or diagnosing a | | |
| CC | cardiovascular, endothelial or angiogenic disorder, including cardiac | | |
| CC | hypertrophy, trauma, cancer, age-related macular degeneration, | | |
| CC | atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis, | | |
| CC | angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour | | |
| CC | angiogenesis (such as breast carcinoma and liver carcinoma) and wound | | |
| CC | healing. The present sequence is a coding sequence of the invention. | | |
| XX | XX | | |
| SQ | Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other; | | |
| Query Match | 98.7%; Score 1337.4; DB 24; Length 2005; | | |
| Best Local Similarity | 99.9%; Pred No. 0; | | |
| Matches 1349; Conservative | 0; Mismatches 1; Indels 1; Gaps 1; | | |
| Qy | 1 | ATGCAACAAGATTTCATGAGGTTCGAGGTAGACACTGAAGTAGCCAACTATACGTGATT | 60 |
| Db | 478 | ATGCAACAAGATTTCATGAGGTTCGAGGTAGACACTGAAGTAGCCAACTATACGTGATT | 537 |
| Qy | 61 | ATGGAAGAATAAGCTAGTACACTCCAAAGCATGCTCAGCTCATCAAGATTTTACAATA | 120 |
| Db | 538 | ATGGAAGAATAAGCTAGTACACTCCAAAGCATGCTCAGCTCATCAAGATTTTACAATA | 597 |
| Qy | 121 | CTACAAGGTCCACCGGCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCAGGACCCCT | 180 |
| Db | 598 | CTACAAGGTCCACCGGCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCAGGACCCCT | 657 |
| Qy | 181 | GGCCCACTGGCAACAAGGGACAGAAAGAGAGAGAGGGGAGCCTGGACACCTGGCCCT | 240 |
| Db | 658 | GGCCCACTGGCAACAAGGGACAGAAAGAGAGAGAGGGGAGCCTGGACACCTGGCCCT | 717 |
| Qy | 241 | CGGGGTGAGAGAGGCCCAATTGGACCACTGGTTCCTCCCGGAGAGCGTGGCGGCAAGA | 300 |
| Db | 718 | CGGGGTGAGAGAGGCCCAATTGGACCACTGGTTCCTCCCGGAGAGCGTGGCGGCAAGA | 777 |
| Qy | 301 | TCTAAAGGCTCCAGGGGCCCAAAAGGCTCCCGTGGTTCCTCCCGGAGAGCGTGGCGG | 360 |
| Db | 778 | TCTAAAGGCTCCAGGGGCCCAAAAGGCTCCCGTGGTTCCTCCCGGAGAGCGTGGCGG | 837 |
| Qy | 361 | GGCCCCACTGGGGACCCAGGCCCCCGGGGCCCAACAGGCAAGAGGGACTTCCCGGCCCT | 420 |
| Db | 838 | GGCCCCACTGGGGACCCAGGCCCCCGGGGCCCAACAGGCAAGAGGGACTTCCCGGCCCT | 897 |
| Qy | 421 | CAGGGCCCTCTGGCTTCAGGGACTTCAGGGACCTTGGGGACCTGGGGTGCCTGGA | 480 |
| Db | 898 | CAGGGCCCTCTGGCTTCAGGGACTTCAGGGACCTTGGGGACCTGGGGTGCCTGGA | 957 |
| Qy | 481 | CCTCGGGGACTGCCAGGCTTGCTGGGGTACCAGGCATGCCAGGCCCAAGAGGCCCCCT | 540 |
| Db | 958 | CCTCGGGGACTGCCAGGCTTGCTGGGGTACCAGGCATGCCAGGCCCAAGAGGCCCCCT | 1017 |
| Qy | 541 | GGCCCTCTGTGCCCCATCAGGAGCGGTGTGGCCCTTGCAGATGAGCCCAACCCCG | 600 |

CC chromosome and gene mapping. ABL88259 to ABL88267 represent primers and
XX probes used in the exemplification of the present invention.

SQ Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;

Query Match 98.7%; Score 1337.4; DB 24; Length 2005;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGCAACAGATTTCATGAGTTCGAGGTAGACACTGAAGTACCAACTTATCAGTGATT 60
DB 478 ATGCAACAGATTTCATGAGTTCGAGGTAGACACTGAAGTACCAACTTATCAGTGATT 537
QY 61 ATGGAAGAAATGAAGTAGTAGACTCCAGCATGGTTCAGCTCATCAAGAATTTTACAATA 120
DB 538 ATGGAAGAAATGAAGTAGTAGACTCCAGCATGGTTCAGCTCATCAAGAATTTTACAATA 597
QY 121 CTACAGAGTCCACCGGGCCCCAGGGTCCAGAGGTGACAGAGGATCCAGGAGCCCGCT 180
DB 598 CTACAGAGTCCACCGGGCCCCAGGGTCCAGAGGTGACAGAGGATCCAGGAGCCCGCT 657
QY 181 GCGCCAACTGGCAACAGGACAGAAAGGAGAGAGGGAGGCCCTGGACCACTGGCCCT 240
DB 658 GCGCCAACTGGCAACAGGACAGAAAGGAGAGAGGGAGGCCCTGGACCACTGGCCCT 717
QY 241 GCGGGTGAGAGAGGCCCAATTGGACAGCTGGTCCCGCCCGGAGAGCGTGGCGGCAAGGA 300
DB 718 GCGGGTGAGAGAGGCCCAATTGGACAGCTGGTCCCGCCCGGAGAGCGTGGCGGCAAGGA 777
QY 301 TCTAAGAGTCCACGGGGCCCCAAGAGCTCCCGTGGTTCCTCGGGAAGCCCGGCTCAG 360
DB 778 TCTAAGAGTCCACGGGGCCCCAAGAGCTCCCGTGGTTCCTCGGGAAGCCCGGCTCAG 837
QY 361 GCGCCCACTGGGAGCCAGGCCCCCGGCCCCCAGGAGGAGGACTCCCGGCGCCT 420
DB 838 GCGCCCACTGGGAGCCAGGCCCCCGGCCCCCAGGAGGAGGACTCCCGGCGCCT 897
QY 421 CAGGGCCCTCCCTGGGTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCCTGGGTGCTGGA 480
DB 898 CAGGGCCCTCCCTGGGTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCCTGGGTGCTGGA 957
QY 481 CTTGGGGAGTCCAGGCTTGCTGGGTACCAAGCATGCCAGGCCCCCAAGGGCCCCCCC 540
DB 958 CTTGGGGAGTCCAGGCTTGCTGGGTACCAAGCATGCCAGGCCCCCAAGGGCCCCCCC 1017
QY 541 GCGCCCTCCCTGGGCCCCATCAGGAGCGGTGGTGGCCCTGGCCCTCAGAAAGCAACCCCG 600
DB 1018 GCGCCCTCCCTGGGCCCCATCAGGAGCGGTGGTGGCCCTGGCCCTCAGAAAGCAACCCCG 1077
QY 601 GCACGGAGGACAATGGCTGCGCCGCTCACTGGAGAGACTTCACAGACAAATGCTACTAT 660
DB 1078 GCACGGAGGACAATGGCTGCGCCGCTCACTGGAGAGACTTCACAGACAAATGCTACTAT 1137
QY 661 TTTTCAGTTGAGAAAGAAATTTTGGAGTGAAGACTTTTCTGCAAGACAAGTCTTCA 720
DB 1138 TTTTCAGTTGAGAAAGAAATTTTGGAGTGAAGACTTTTCTGTAAGACAAGTCTTCA 1197
QY 721 CATCTTGTGTTTTCATAAAGTACAGAGGAAACAGCAATGGATAAAAAACAGATGGTAGG 780
DB 1198 CATCTTGTGTTTTCATAAAGTACAGAGGAAACAGCAATGGATAAAAAACAGATGGTAGG 1257
QY 781 AGAGAGAGCCACTGGATGGCCCTCAGACTCAGAGCGTGAATAATGAATGGAAGTGGGTG 840
DB 1258 AGAGAGAGCCACTGGATGGCCCTCAGACTCAGAGCGTGAATAATGAATGGAAGTGGGTG 1317
QY 841 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCCGGATACCTGGGTGAT 900
DB 1318 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCCGGATACCTGGGTGAT 1377
QY 901 GGCCATGGGCCAGGAGAGACTGTGCTGGGTTGATTTATGCTGGGAGTGAAGCAATTC 960
DB 1378 GGCCATGGGCCAGGAGAGACTGTGCTGGGTTGATTTATGCTGGGAGTGAAGCAATTC 1437

QY 961 CAATGTGAAGAGCGTCAATAAATCTTATTTGCGAAAAAGACAGGGAGACAGTACTGTCTATCT 1020
DB 1438 CAATGTGAAGAGCGTCAATAAATCTTATTTGCGAAAAAGACAGGGAGACAGTACTGTCTATCT 1497
QY 1021 GCATTATACCGGACTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAGAGGCAAGG 1080
DB 1498 GCATTATACCGGACTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAGAGGCAAGG 1557
QY 1081 ACACATCCCTTTCTAATTTGATCATCACCTTCTCATCAGATTGAAAAAAGAACGACTGAAA 1140
DB 1558 ACACATCCCTTTCTAATTTGATCATCACCTTCTCATCAGATTGAAAAAAGAACGACTGAAA 1616
QY 1141 CCAATTACTGAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC 1200
DB 1617 CCAATTACTGAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC 1676
QY 1201 TTGGGAACATAAATGTTCCCGAGGGTGATATGCTGATTTTCATTGTGCACATGGACTGAA 1260
DB 1677 TTGGGAACATAAATGTTCCCGAGGGTGATATGCTGATTTTCATTGTGCACATGGACTGAA 1736
QY 1261 TCACATAGATTCTCTCCGTCAGTAACCGTGGGATTATACAAATTTATGCTTCCAAAGTA 1320
DB 1737 TCACATAGATTCTCTCCGTCAGTAACCGTGGGATTATACAAATTTATGCTTCCAAAGTA 1796
QY 1321 TGGAACTCCCAATCAGAAAAAGGTTATCAT 1351
DB 1797 TGGAACTCCCAATCAGAAAAAGGTTATCAT 1827

RESULT 5

AA158842

ID AA158842 standard; cDNA; 2318 BP.

XX AC AA158842;

XX XX 22-OCT-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 1045.

XX KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia; ss.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, xue AJ, yang Y, Zhang J;

XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI; 2001-442253/47.

XX DR P-PSDB; AAM39686.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders

| | |
|---|---|
| PT | such as central nervous system injuries - |
| XX | |
| PS | Claim 1; SEQ ID NO 1045; 10078pp; English. |
| XX | |
| CC | The invention relates to human nucleic acids (AA157798-AA161369) and |
| CC | the encoded polypeptides (AA138642-AA142213) with nootropic, |
| CC | immunosuppressant and cytostatic activity. The polynucleotides are useful |
| CC | in gene therapy. A composition containing a polypeptide or polynucleotide |
| CC | of the invention may be used to treat diseases of the peripheral nervous |
| CC | system, such as peripheral nervous injuries, peripheral neuropathy and |
| CC | localised neuropathies and central nervous system diseases, such as |
| CC | Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic |
| CC | lateral sclerosis, and Shy-Drager Syndrome. Other uses include the |
| CC | utilisation of the activities such as: immune system suppression, |
| CC | Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic |
| CC | and thrombolytic activity, cancer diagnosis and therapy, drug screening, |
| CC | assays for receptor activity, arthritis and inflammation, leukaemias and |
| CC | C.N.S disorders. |
| CC | Note: The sequence data for this patent did not form part of the printed |
| CC | specification. |
| XX | |
| SEQ | Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other; |
| Query Match 98.7%; Score 1337.4; DB 22; Length 2318; | |
| Best Local Similarity 99.9%; Pred. No. 0; | |
| Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1; | |
| QY | 1 ATGCAACAGATTGATGAGGTCGAGGTAGACACTGAAGTAGCAACTTATCAGTGATT 60 |
| DB | 663 ATGCAACAGATTGATGAGGTCGAGGTAGACACTGAAGTAGCAACTTATCAGTGATT 722 |
| QY | 61 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGGATGTCAGCTCATCAAGAATTTTACAATA 120 |
| DB | 723 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTCAGCTCATCAGAATTTTACAATA 782 |
| QY | 121 CTAAAGGTTCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCCAGGGACCCCTT 180 |
| DB | 783 CTAAAGGTTCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCCAGGGACCCCTT 842 |
| QY | 181 GGCCCAACTGGCAACAAGGCACAGAAGAGAGAGGGGAGCGCTGGACCACTGGCCCT 240 |
| DB | 843 GGCCCAACTGGCAACAAGGCACAGAAGAGAGAGGGGAGCGCTGGACCACTGGCCCT 902 |
| QY | 241 CGGGGTGAGAGAGGCCCAATTGGTACACAGCTGGTCCCCCGAGAGCGTGGCGCAAGGA 300 |
| DB | 903 CGGGGTGAGAGAGGCCCAATTGGTACACAGCTGGTCCCCCGAGAGCGTGGCGCAAGGA 962 |
| QY | 301 TCTAAAGGCTCCAGGGCCCCAAAGGCTCCGCTGGTTCCTTGGGAAGCCCGGCCCTTCAG 360 |
| DB | 963 TCTAAAGGCTCCAGGGCCCCAAAGGCTCCGCTGGTTCCTTGGGAAGCCCGGCCCTTCAG 1022 |
| QY | 361 GGCCCCAGTGGGACCCAGGCCCCCCGGGCCCCACAGCAAGAGAGGAGCTCCCGGGCCCT 420 |
| DB | 1023 GGCCCCAGTGGGACCCAGGCCCCCCGGGCCCCACAGCAAGAGAGGAGCTCCCGGGCCCT 1082 |
| QY | 421 CAGGGCCCTCCTGGCTTCCAGGGACTTCAGGGCAACCGTTGGGGAGCGCTGGGGTGCCTGGA 480 |
| DB | 1083 CAGGGCCCTCCTGGCTTCCAGGGACTTCAGGGCAACCGTTGGGGAGCGCTGGGGTGCCTGGA 1142 |
| QY | 481 CTTGGGGAGACTGCAGGCTTTGGCTTGGGTATCAGGCAATGCAGGCCCCCAAGGGCCCCCCC 540 |
| DB | 1143 CTTGGGGAGACTGCCAGGCTTTGGCTTGGGTATCAGGCAATGCAGGCCCCCAAGGGCCCCCCC 1202 |
| QY | 541 GGCCCTTCCTGGCCCATCAGGAGCGGTGTGCCCTTGGCCCTGCAGATGAGCCACACCCCG 600 |
| DB | 1203 GGCCCTTCCTGGCCCATCAGGAGCGGTGTGCCCTTGGCCCTGCAGATGAGCCACACCCCG 1262 |
| QY | 601 GCACCGGAGCAATGGGCTGCGCGCTCAGCTGGGAAGAACTTCACAGACAATGCTACTAT 660 |
| DB | 1263 GCACCGGAGCAATGGGCTGCGCGCTCAGCTGGGAAGAACTTCACAGACAATGCTACTAT 1322 |
| QY | 661 TTTTCAGTTGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAGTCTTCA 720 |

PA (INCY-) INCYTE GENOMICS INC.

XX
PI Bandman O, Hillman JL, Tang YT, Yue H, Baughn MR, Lu DM;
PI Azimzai Y;
XX
DR WPI; 2001-007395/01.
DR P-PSDB: AAB27236.

Isolated polynucleotide encoding extracellular matrix or adhesion-associated protein (EXMAD) useful for diagnosing, treating, or preventing disorders associated with expression of EXMAD such as proliferative, immune and genetic disorders -

PS Claim 4; page 121-122: 129pp: English.

The present invention provides the protein and coding sequences for 25 these extracellular matrix and adhesion-associated proteins (EXMADS). These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5, EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12, EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19, EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are useful in the prevention and treatment of cancers, cell proliferation, cardiovascular, reproductive, immune, musculoskeletal, developmental and gastrointestinal disorders and inflammation.

SQ Sequence 2641 BP; 823 A; 628 C; 608 G; 582 T; 0 other:

| | | | | |
|-----------------------|--------------|--------------|------------|-------------|
| Query Match | 98.7% | Score 1337.4 | DB 22 | Length 2641 |
| Best Local Similarity | 99.9% | Pred. No. 0 | | |
| Matches 1349 | Conservative | 0 | Mismatches | 1 |
| | | | Indels | 1 |
| | | | Gaps | 1 |

1 ATGCAACAAGATTTGATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 60

Db 912 ATGCAACAAGATTTCATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGAATT 971

61 ATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAAATTTACAATA 120

db 972 ATCGAAGAAATCAAGCTACTACCTCACAACCATTCCCTCATTCATTA 1021

100

1033

nk

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

DW 1Z1Z 1C1AAG0C1CCCA000GCCCCCAAAG0C1CCCIG0I1CCCCI0GGAAAGCCCCICAG 1Z7T

QY J01 90CCCCAGT0000HCCCCAGGGCCCCCGGGGCCACCGGGCAAHAGAGGGAC1CCCCGGCCCT 420

DD 12/2 GGCCCCAGTGGGGACCCAGGCCCCCGGGCCACAGGCAAGAGGGACTCCCCGGCCCT 1331

QY 421 CAGGGCCCTGCCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCTTGGGGTGCCCTGGA 480

db 1332 CAGGGCCCTCCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCCTGGGGTGCCCTGGA 1391

QY 481 CCTCGGGGACTGCCAGGCTTGCCCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCCC 540

Db 1392 CCTCGGGACTGCCAGGCTTGCCCTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCC 1451

QY 541 GGCCCTCCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCGAATGAGCCAACCCCG 600

Db 1452 GGCCCTCCTGGCCCATCAGGAGCGTGGTGGCCCTGGCCCTGCAGATGAGCCAACCCCG 1511

QY 601 GCACCGAGGACAAATGGCTGCCCGCCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 660

| | | | | |
|-----------------------------------|-----|--|---|------|
| D | b | 1512 | GCACCGGAGGACAATAGCTGCGCGGCTCACTGGGAAGAACTTCACAGACAAATGCTACTAT | 1571 |
| Q | y | 661 | TTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTCA | 720 |
| D | b | 1572 | TTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTCA | 1631 |
| Q | y | 721 | CATCTGTTTTCATAAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGG | 780 |
| D | b | 1632 | CATCTGTTTTCATAAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGG | 1691 |
| Q | y | 781 | AGAGAGAGCCACTGGATCGGCCTCAGAGACTCAGAGCGTGAATGAATGAAGTGGCTG | 840 |
| D | b | 1692 | AGAGAGAGCCACTGGATCGGCCTCAGAGACTCAGAGCGTGAATGAATGAAGTGGCTG | 1751 |
| Q | y | 841 | GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCGGATAACTGGGGTCAT | 900 |
| D | b | 1752 | GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCGGATAACTGGGGTCAT | 1811 |
| Q | y | 901 | GGCCATGGGCGAGGAAGACTGTGCTGGGTTGATTTATGCTGGGCACTGGAACGATTTTC | 960 |
| D | b | 1812 | GGCCATGGGCGAGGAAGACTGTGCTGGGTTGATTTATGCTGGGCACTGGAACGATTTTC | 1871 |
| Q | y | 961 | CAATGTGAAGACGTCATTAACITTCATTTGCGAAAAGACGGGAGACAGTACTGTCATCT | 1020 |
| D | b | 1872 | CAATGTGAAGACGTCATTAACITTCATTTGCGAAAAGACGGGAGACAGTACTGTCATCT | 1931 |
| Q | y | 1021 | GCATTATAACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGCTCTCAAAAGCAAAAG | 1080 |
| D | b | 1932 | GCATTATAACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGCTCTCAAAAGCAAAAG | 1991 |
| Q | y | 1081 | ACACTCCTTTCTAATTGCAATCACTTCTCATCAGATTTGAAAAAAGCACTGAAAA | 1140 |
| D | b | 1992 | ACACTCCTTTCTAATTGCAATCACTTCTCATCAGATTTGAAAAAAGCACTGAAAA | 2050 |
| Q | y | 1141 | CCAACTACTCAAAAAAATTTGACACTAGTGTTTTTTACCATCCGTCATTAACCCAAAGAC | 1200 |
| D | b | 2051 | CCAACTACTCAAAAAAATTTGACACTAGTGTTTTTTACCATCCGTCATTAACCCAAAGAC | 2110 |
| Q | y | 1201 | TTGGGAACTAAAATGTTCCCGAGGTGATATGCTGATTTTCATTTGTCACATGGACTGAA | 1260 |
| D | b | 2111 | TTGGGAACTAAAATGTTCCCGAGGTGATATGCTGATTTTCATTTGTCACATGGACTGAA | 2170 |
| Q | y | 1261 | TCACATAGATTCTCCCTCGTCACTAACCGTGGGATTTATACAAATATGCTTTCCAAAGTA | 1320 |
| D | b | 2171 | TCACATAGATTCTCCCTCGTCACTAACCGTGGGATTTATACAAATATGCTTTCCAAAGTA | 2230 |
| Q | y | 1321 | TGGAACACTCCATCATGAAAAAGGTTATCAT | 1351 |
| D | b | 2231 | TGGAACACTCCATCATGAAAAAGGTTATCAT | 2261 |
| RESULT 7 | | | | |
| AAV55746 standard; cdNA; 2930 BP. | | | | |
| XX | AC | AAV55746; | | |
| XX | XX | 23-MAR-1999 (first entry) | | |
| XX | XX | Human secreted protein clone bv227_1 coding sequence. | | |
| DE | DE | Secreted protein; human; nutritional supplements; immune stimulant; | | |
| KW | KW | immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis | | |
| KW | KW | activin/inhibin; chemokinesis; haemostasis; thrombolysis; | | |
| KW | KW | receptor/ligand activity; anti-inflammatory; tumour inhibitor; | | |
| XX | XX | cadherin/tumour invasion suppressor; ds. | | |
| OS | OS | Homo sapiens. | | |
| XX | XX | | | |
| FH | Key | Location/Qualifiers | | |
| FT | CDS | 67..693 | | |
| FT | FT | /*tag= a | | |
| XX | XX | | | |

RESULT 7

AAV55746

ID AAV55746 standard; cDNA; 2930 BP.

AC AAV55746;

DT 23-MAR-1999 (first entry)

DE Human secreted protein clone bv227_1 coding sequence.

KW Secreted protein; human; nutritional supplements; immune stimulant;
 KW Immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis
 KW activin/inhibin; chemokinesis; haemostasis; thrombolysis;
 KW receptor/ligand activity; anti-inflammatory; tumour inhibitor;
 KW cadherin/tumour invasion suppressor; ds.
 KW

AA Homo sapiens.

| XX | Key | Location/Qualifiers |
|----|-----|---------------------|
| EH | | |

671.093
/ * + 3 3 3
CDS

XX

PN W09855614-A2.
XX 10-DEC-1998.
XX 01-JUN-1998; 98WO-US11210.
XX 29-MAY-1998; 98US-0087255.
PR 04-JUN-1997; 97US-0868696.
PR 04-JUN-1997; 97US-0868697.
PR 04-JUN-1997; 97US-0868698.
PR 04-JUN-1997; 97US-0868898.
PR 04-JUN-1997; 97US-0868899.
PR 04-JUN-1997; 97US-0868900.
PR 04-JUN-1997; 97US-0869191.
PR 04-JUN-1997; 97US-0869192.
PR 04-JUN-1997; 97US-0869193.
PR 04-JUN-1997; 97US-0869194.
XX (GENY) GENETICS INST INC.
XX Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER;
PI McCoy JM, Racie LA, Spaulding V, Treacy M;
XX WPI: 1999-059912/05.
DR P-PSDB; AAW73628.
XX New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult brain, foetal kidney, adult ovary, adult
PT retina, adult placenta or adult uterus cDNA libraries
XX Claim 26; Page 87-88; 127pp; English.
XX This sequence encodes a human secreted protein of the invention.
CC This DNA sequence was isolated from a human adult brain cDNA
CC library, and was designated clone bv227.1. The DNAs and proteins
CC are predicted to have biological activities which would make them
CC suitable for treating, preventing or ameliorating medical conditions in
CC humans and animals, although no supporting data is given. Suggested
CC activities include nutritional sources or supplements, immune
CC stimulating or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity. The DNAs
CC are also stated to be useful for gene therapy. A host cell transfected
CC with the DNA, or its subfragments and variants is useful for recombinant
CC production of the human secreted protein clones.
XX
SQ Sequence 2930 BP; 902 A; 699 C; 692 G; 537 T; 0 other;
Query Match 98.6%; Score 1335.8; DB 20; Length 2930;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1348; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 ATGCACACAGATTTCATGAGTTCGAGGTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
DB 1266 ATGCCAACAGATTTCATGAGTTCGAGGTAGACACTGAAGTAGCCAACTTATCAGTGATT 1325
QY 61 ATGGGAAGAAATCAAGCTAGTACCTCCAGCATGTGTGAGTCTCATCAAGAATTTTACAATA 120
DB 1326 ATGGGAAGAAATGAGCTAGTACCTCCAGCATGTGTGAGTCTCATCAAGAATTTTACAATA 1385
QY 121 CTACAAGTTCACCGGGCCCCAGGGTCCAGAGGTGACAGAGGATCCACAGGACCCCT 180
DB 1386 CTACAAGTTCACCGGGCCCCAGGGTCCAGAGGTGACAGAGGATCCACAGGACCCCT 1445
QY 181 GSCCAACTGGCAACAAGGACAGAAAGAGAGAGGGGGAGCCCTGGACCCTGGCCCT 240
DB 1446 GSCCAACTGGCAACAAGGACAGAAAGAGAGAGGGGGAGCCCTGGACCCTGGCCCT 1505
QY 241 GCGGGTGAGAGAGGCCCAATTGACACAGCTGTGCTCCCGGAGAGCGTGGCGGCAAGGA 300
DB 1506 GCGGGTGAGAGAGGCCCAATTGACACAGCTGTGCTCCCGGAGAGCGTGGCGGCAAGGA 1565

QY 301 TCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAG 360
DB 1566 TCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAG 1625
QY 361 GCGCCAGTGGGGAGCCAGGCCCCCGGGCCACACAGGAAAGAGGACTCCCGGCCCT 420
DB 1626 GCGCCAGTGGGGAGCCAGGCCCCCGGGCCACACAGGAAAGAGGACTCCCGGCCCT 1685
QY 421 CAGGGCCCTCCCTGGCTCCAGGGACTTCAGGACACCGTTGGGAGGCTGGGTGCTTGA 480
DB 1686 CAGGGCCCTCCCTGGCTTCAGGGACTTCAGGGACCGTTGGGAGGCTGGGTGCTTGA 1745
QY 481 CCTCGGGACTGCCAGGCTTGCTGGGTACCAAGGATGCCAGGCCCAAGAGGCCCCCTCC 540
DB 1746 CCTCGGGACTGCCAGGCTTGCTGGGTACCAAGGATGCCAGGCCCAAGAGGCCCCCTCC 1805
QY 541 GCGCCTCTCGGCCCATCAGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCG 600
DB 1806 GCGCCTCTCGGCCCATCAGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCG 1865
QY 601 GCACCGGAGCAATGGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 660
DB 1866 GCACCGGAGCAATAGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 1925
QY 661 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAAGTCTTCA 720
DB 1926 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAAGTCTTCA 1985
QY 721 CATCTGTTTTCATAAACACTAGAGAGAACAGCAATGGATAAATAACAGATGTTAGG 780
DB 1986 CATCTGTTTTCATAAACACTAGAGAGAACAGCAATGGATAAATAACAGATGTTAGG 2045
QY 781 ACAGAGAGCACTGGATCGGCTCCAGACTCAGAGCGTGAAGTGAAGTGGCTG 840
DB 2046 AGAGAGAGCACTGGATCGGCTCCAGACTCAGAGCGTGAAGTGAAGTGGCTG 2105
QY 841 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGATCACTGGGTCAT 900
DB 2106 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGATCACTGGGTCAT 2165
QY 901 GGCATGGCCAGGAGAGAACTGTGCTGGTTGATTTATGCTGGGAGTGGACGATTTC 960
DB 2166 GGCATGGCCAGGAGAGAACTGTGCTGGTTGATTTATGCTGGGAGTGGACGATTTC 2225
QY 961 CAATGTGAAGAGCTCAATAACTTTCGAAAAGACAGGAGACAGTACTGTCTATCT 1020
DB 2226 CAATGTGAAGAGCTCAATAACTTTCGAAAAGACAGGAGACAGTACTGTCTATCT 2285
QY 1021 GCATTATAACGAGCTGTGATGGATCAGATGACAAATTTTCACTCTCAAGGCAAGG 1080
DB 2286 GCATTATAACGAGCTGTGATGGATCAGATGACAAATTTTCACTCTCAAGGCAAGG 2345
QY 1081 ACACCTCTTCTTAATTCATCCTTCTCATCAGATTGAAAAAAGACACTGAAAA 1140
DB 2346 ACACCTCTTCTTAATTCATCCTTCTCATCAGATTGAAAAAAGACACTGAAAA 2404
QY 1141 CCAATTACTGAAAAAATTTGACAGTACTGTTTTTACCATCCGTCATTACCCAAAGAC 1200
DB 2405 CCAATTACTGAAAAAATTTGACAGTACTGTTTTTACCATCCGTCATTACCCAAAGAC 2464
QY 1201 TTGGGAACATAAATGTTCCCGAGGGTGATGCTGATTTTTCATTTGTCACATGGACTGAA 1260
DB 2465 TTGGGAACATAAATGTTCCCGAGGGTGATGCTGATTTTTCATTTGTCACATGGACTGAA 2524
QY 1261 TCACATAGATTCTCCCTCCGTCAGTAACCGTGGATTAACAATTTATGTTCTTCCAAAGTA 1320
DB 2525 TCACATAGATTCTCCCTCCGTCAGTAACCGTGGATTAACAATTTATGTTCTTCCAAAGTA 2584
QY 1321 TGAACACACTCCCAATCAGAAAAAGGTTTATCAT 1351
DB 2585 TGAACACACTCCCAATCAGAAAAAGGTTTATCAT 2615

RESULT 8
AAI60628
ID AAI60628 standard; cDNA; 2181 BP.
XX
AC AAI60628;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4617.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
PI WPI: 2001-442253/47.
DR P-PSDB; AAM41472.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 4617; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2181 BP; 664 A; 539 C; 535 G; 443 T; 0 other;

Query Match 98.1%; Score 1328.8; DB 22; Length 2181;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1344; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 1 ATGCACACAGATTGATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTATCAGTGATT 60

Db 662 ATGCACACAGATTGATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTATCAGTGATT 721
QY 61 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCAGTGGTCAGCTCATCAAGAAATTTACAATA 120
Db 722 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCAGTGGTCAGCTCATCAAGAAATTTACAATA 781
QY 121 CTAAAGGTTCCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGGATCCCAAGGACCCCT 180
Db 782 CTAAAGGTTCCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGGATCCCAAGGACCCCT 841
QY 181 GGGCCAACTGGCAACAAAGGACAGAAAGAGAGAAAGGGGAGGCTGGACACCTGGCCCT 240
Db 842 GGGCCAACTGGCAACAAAGGACAGAAAGAGAGAAAGGGGAGGCTGGACACCTGGCCCT 901
QY 241 GCGGGTGAGAGAGGCGCAATTTGGACAGCTGGTCCCGCGGAGAGCGTGGCGCAAGGA 300
Db 902 GCGGGTGAGAGAGGCGCAATTTGGACAGCTGGTCCCGCGGAGAGCGTGGCGCAAGGA 961
QY 301 TCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCGGAAAGCGCCCTCAG 360
Db 962 TCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCGGAAAGCGCCCTCAG 1021
QY 361 GGGCCAGTGGGAGACCCAGGCCCCCGGGCCACAGGCAAGAGGAGCTCCCGGGCCCT 420
Db 1022 GGGCCAGTGGGAGACCCAGGCCCCCGGGCCACAGGCAAGAGGAGCTCCCGGGCCCT 1081
QY 421 CAGGGCCCTCCTGGCTTCAGGGACTTCAGGGCACCCTGGGGAGCGCTGGGGTGCCTGA 480
Db 1082 CAGGGCCCTCCTGGCTTCAGGGACTTCAGGGCACCCTGGGGAGCGCTGGGGTGCCTGA 1141
QY 481 CCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAAGGATGCCAGGCCCCCAAGGCCCCCCC 540
Db 1142 CCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAAGGATGCCAGGCCCCCAAGGCCCCCCC 1201
QY 541 GGGCCCTCCTGGCCCATCAGAGGCGGTGGTCCCTCGCCCTGCAGAAATGAGCCAAACCCCG 600
Db 1202 GGGCCCTCCTGGCCCATCAGAGGCGGTGGTCCCTCGCCCTGCAGAAATGAGCCAAACCCCG 1261
QY 601 GCACCGGAGACAAATGGCTGCCCGCTCCTACTGGAAGAACTTCACAGCAAAATGCTACTAT 660
Db 1262 GCACCGGAGACAAATGCTGCGCCCTCCTACTGGAAGAACTTCACAGCAAAATGCTACTAT 1321
QY 661 TTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGAGCTTTTCTGTGAAGACAAAGTCTTCA 720
Db 1322 TTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGAGCTTTTCTGTGAAGACAAAGTCTTCA 1381
QY 721 CATCTGTTTTCATAAACAAGAGAGGACCAATGGATAAAAAACAGATGGTAGGG 780
Db 1382 CATCTGTTTTCATAAACAAGAGAGGACCAATGGATAAAAAACAGATGGTAGGG 1441
QY 781 AGAGAGAGCCACTGGATCGCCCTCAGAGACTCAGAGCGTGAAATGAATGGAAGTGGCTG 840
Db 1442 AGAGAGAGCCACTGGATCGCCCTCAGAGACTCAGAGCGTGAAATGAATGGAAGTGGCTG 1501
QY 841 GATGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAT 900
Db 1502 GATGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAT 1561
QY 901 GGGCCATGGCCAGGAGAAAGCTGTGCTGGGTGATTATGCTGGGAGTGGAAACGATTTC 960
Db 1562 GGGCCATGGCCAGGAGAAAGCTGTGCTGGGTGATTATGCTGGGAGTGGAAACGATTTC 1621
QY 961 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAAGACAGGAGAGACGTACTGTCTATCT 1020
Db 1622 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAAGACAGGAGAGACGTACTGTCTATCT 1681
QY 1021 GCATTATAACGAGCTGTGATGGGATCACATGAGCAAAATTTTTCAGCTCTCAAGGCAAGG 1080
Db 1682 GCATTATAACGAGCTGTGATGGGATCACATGAGCAAAATTTTTCAGCTCTCAAGGCAAGG 1741
QY 1081 ACACCTCTTTCTTAATTTGATGATCAGCTTCTCATGATTTGAAAAAAGAGGAGCTCAAAA 1140

| | | | |
|----------|--|---|------|
| Db | 1742 | ACACTCTTTCTTAATTCGCATCACACCTTCTCATCATGAGTTGAAAAAAGACACTGAAAG | 1801 |
| Qy | 1141 | CCAATTACTGAAAAAAATTCAGACAGCTAGTGTCTTTTACCATCCGTCAATACCCAAAGAC | 1200 |
| Db | 1802 | CCATTTACTGAAAAAAATTCAGACAGCTAGGTCTTTTACCATCCGTCAATACCCAAAGAC | 1861 |
| Qy | 1201 | TTGGGAACATAAATGTTCCCGCAGGGTGATATGCTGATTTTCATTTGTCACATGCATGAA | 1260 |
| Db | 1862 | TTGGGAACATAAATGTTCCCGCAGGGGATATGCTGATTTTCATTTGGGCACATGCATGAA | 1921 |
| Qy | 1261 | TCACATAGATTCTCCTCCGTCAGTAACCGT-GGATTATACAAATATATGCTTTCCCAAGT | 1319 |
| Db | 1922 | TCACATAAATCTCCTCCGTCAGTAACCGTGGCGATTATACAAATATATGCTTTCCCAAGT | 1981 |
| Qy | 1320 | ATGGAACACTCCCAATCAGAAAAAGTTATCAT | 1351 |
| Db | 1982 | ATGGAACACTCCCAATCAGAAAAAGTTATCAT | 2013 |
| RESULT 9 | | | |
| ID | ABQ92072 | | |
| XX | ABQ92072 standard; cDNA; 2929 BP. | | |
| XX | ABQ92072; | | |
| XX | 04-OCT-2002 (first entry) | | |
| XX | Human polynucleotide SEQ ID NO 69. | | |
| DE | | | |
| XX | Human; cytostatic; antirheumatic; antiarthritic; vulnery; analgesic; | | |
| KW | antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian; | | |
| KW | neuroprotective; nootropic; osteopathic; haemostatic; vasotropic; | | |
| KW | antitumor; fungicide; antidiabetic; antiasthmatic; antiallergic; | | |
| KW | immunostimulant; antiparasitic; secreted protein; transmembrane protein; | | |
| KW | cytokine; cell proliferation; cell differentiation; autoimmune disease; | | |
| KW | stem cell; growth factor; nervous system disease; neuropathy; | | |
| KW | Alzheimer's disease; Parkinson's disease; Huntington's disease; | | |
| KW | osteoporosis; severe combined immunodeficiency; SCID; infection; | | |
| KW | multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss. | | |
| XX | | | |
| OS | homo sapiens. | | |
| XX | | | |
| PN | US2002065394-A1. | | |
| XX | | | |
| PD | 30-MAY-2002. | | |
| XX | | | |
| PF | 22-DEC-2000; 2000US-0745763. | | |
| XX | | | |
| PR | 18-MAR-1998; 98US-0040963. | | |
| XX | | | |
| PA | (JACO/) JACOBS K. | | |
| PA | (MCCO/) MCCOY J M. | | |
| PA | (LAVA/) LAVALLIE E R. | | |
| PA | (COLL/) COLLINS-RACIE L A. | | |
| PA | (EVAN/) EVANS C. | | |
| PA | (MERB/) MERBERG D. | | |
| PA | (TREA/) TREACY M. | | |
| PA | (SPAU/) SPAULDING V. | | |
| XX | | | |
| PI | Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C; | | |
| PI | Merberg D, Treacy M, Spaulding V; | | |
| XX | | | |
| XX | WPI: 2002-582343/62. | | |
| DR | P-PSDB; ABP61859. | | |
| XX | | | |
| PT | Novel secreted or transmembrane protein and polynucleotide encoding the | | |
| PT | protein, useful for diagnosis and treatment of neurological disorders, | | |
| PT | cancer, autoimmune diseases, bone disorders and lung or liver fibrosis | | |
| XX | | | |
| PS | Claim 250; Page 231-232; 284pp; English. | | |
| XX | | | |
| CC | The invention relates to human secreted or transmembrane protein (I). | | |

CC their fragments and is encoded by specific complementary deoxyribonucleic
 CC acid (cDNA) inserts (11), where the protein is substantially free from
 CC other mammalian proteins. (1) are useful for preventing, treating or
 CC ameliorating a medical condition, especially immunological treatment or
 CC prevention of tumours. (1) exhibits activity relating to angiogenesis,
 CC cytokine, cell proliferation, cell differentiation, antiinflammatory,
 CC stem cell growth factor activity and activin or inhibin-related
 CC activities. (1) can be used to manipulate stem cells in culture to give
 CC rise to neuroepithelial cells that can be used to augment or replace
 CC cells damaged by illness, autoimmune disease, accidental damage or
 CC genetic disorders. (1) induces the proliferation of neural cells and
 CC regeneration of nerve and brain tissue and is useful for the treatment of
 CC central and peripheral nervous system diseases and neuropathies, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis. (1) is involved in chemotactic or chemokinetic
 CC activity, regulation of haematopoiesis and is useful for treating myeloid
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopaenia
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
 CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,
 CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
 CC periodontal disease. (1) is also useful for gut protection or
 CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
 CC in various tissues, various immune deficiencies and disorders including
 CC severe combined immunodeficiency (SCID), bacterial or fungal infections,
 CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
 CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
 CC such as asthma or other respiratory problems. (1) is useful to express
 CC recombinant protein, as markers for tissues in which the corresponding
 CC protein is preferentially expressed and in gene therapy. The present
 CC sequence is that of a polynucleotide of the invention.
 CC
 XX
 SQ Sequence 2929 BP; 901 A; 599 C; 692 G; 637 T; 0 other;

 Query Match 97.7%; Score 1323.8; DB 24; Length 2929;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1347; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

 QY 1 ATGCAACAAGATTTCATCAGGTCGAGGTAGACACTGAAGTAGCCCACTTATCAGTGATT 60
 Db |||||
 1266 ATGCAACAAGATTTCATGAGGTCGAGGTAGACACTGAAGTAGCCCACTTATCAGTGATT 1325

 QY 61 ATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAAATTTACAAATA 120
 Db |||||
 1326 ATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAAATTTACAAATA 1385

 QY 121 CTACAAGTCCACCGGGCCCGAGGGTCCAGAGAGGTGACAGAGGATCCAGGGACCCCT 180
 Db |||||
 1386 CTACAAGTCCACCGGGCCCGAGGGTCCAGAGAGGTGACAGAGGATCCAGGGACCCCT 1445

 QY 181 GGCCCAACTGGCAACAAGGGACAGAAAGAGAGGGGAGCGCTGGACACACTGGCCCT 240
 Db |||||
 1446 GGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGGGGAGCGCTGGACACACTGGCCCT 1505

 QY 241 GCGGGTGAGAGAGGCCCAATTTGGACACAGCTGGTCCCGCCGAGAGCGGTGGCGCAAGAAGA 300
 Db |||||
 1506 GCGGGTGAGAGAGGCCCAATTTGGACACAGCTGGTCCCGCCGAGAGCGGTGGCGCAAGAAGA 1565

 QY 301 TCTAAGAGCTCCAGGGCCCAAGAGCTCCCGTGGTTCCCTCGGAGAGCCCGGCCCTCAG 360
 Db |||||
 1566 TCTAAGAGCTCCAGGGCCCAAGAGCTCCCGTGGTTCCCTCGGAGAGCCCGGCCCTCAG 1625

 QY 361 GGCCCCAGTGGGGACCCAGGCCCCCGGGCCCCACAGCAAGAGGACTCCCGGCCCT 420
 Db |||||
 1626 GGCCCCAGTGGGGACCCAGGCCCCCGGGCCCCACAGCAAGAGGACTCCCGGCCCT 1685

 QY 421 CAGGGCCCTCCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCGCTGGGGTGCCCTGGA 480
 Db |||||
 1686 CAGGGCCCTCCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCGCTGGGGTGCCCTGGA 1745

 QY 481 CCTCGGGAGCTGCCAGGCTTGCTCGTGGGTACCAAGGCATGCCAGGCCCAAGGGCCCCCCC 540
 Db |||||
 1746 CCTCGGGAGCTGCCAGGCTTGCTCGTGGGTACCAAGGCATGCCAGGCCCAAGGGCCCCCCC 1805

QY 541 GGCCCTCTCGGCCCATCAGAGCGGTGGTGGCCCTGCAGAAATGAGCAACCCCG 600
Db 1806 GGCCCTCTCGGCCCATCAGAGCGGTGGTGGCCCTGCAGAAATGAGCAACCCCG 1865
QY 601 GCACCGGAGGACAATGGCTGCGCCCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 560
Db 1866 GCACCGGAGGACAATGAGCTGCGCCCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 1925
QY 661 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAAAGCTTTCTGTGAACACAAGTCTTCA 720
Db 1926 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAAAGCTTTCTGTGAACACAAGTCTTCA 1985
QY 721 CATCTGTTTTTCATAAACAATAGAGAGGAACAGCAATGATGATAAAAAACAGATGGTAGGG 780
Db 1986 CATCTGTTTTTCATAAACAATAGAGAGGAACAGCAATGATGATAAAAAACAGATGGTAGGG 2045
QY 781 AGAGAGAGCCACTGGATCGGCCTCACAGACTCAGACGCTGAAATGAATGGAAGTGGCTG 840
Db 2046 AGAGAGAGCCACTGGATCGGCCTCACAGACTCAGACGCTGAAATGAATGGAAGTGGCTG 2105
QY 841 GATGGACATCTCCAGACTACAAAATTTGAAAGCTGGACGCGGATAACTGGGTCTAT 900
Db 2106 GATGGACATCTCCAGACTACAAAATTTGAAAGCTGGACGCGGATAACTGGGTCTAT 2165
QY 901 GGCCATGGGCCAGGAGAAAGACTGTGCTGGGTTGATTTATGCTGGCAGTGAACGATTTTC 960
Db 2166 GGCCATGGGCCAGGAGAAAGACTGTGCTGGGTTGATTTATGCTGGCAGTGAACGATTTTC 2225
QY 961 CAATGTGAAGAGCTCAATTAACCTTCATTTCCGAAAAGAGACAGGACAGTACTGTCACT 1020
Db 2226 CAATGTGAAGAGCTCAATTAACCTTCATTTCCGAAAAGAGACAGGACAGTACTGTCACT 2285
QY 1021 GCATTATAACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGCTCTCAAAAGGCAAAAG 1080
Db 2286 GCATTATAACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGCTCTCAAAAGGCAAAAG 2345
QY 1081 ACATCTCTTTCTAATGTCAATACCTTCTCATCAGATGAAAAAAGCACTGAATA 1140
Db 2346 ACATCTCTTTCTAATGTCAATACCTTCTCATCAGATGAAAAAAGCACTGAATA 2404
QY 1141 CCAATTTACTGAAAAAATTTGACAGTAGTGTGTTTACCATCCGTCATTACCCAAAGAC 1200
Db 2405 CCAATTTACTGAAAAAATTTGACAGTAGTGTGTTTACCATCCGTCATTACCCAAAGAC 2464
QY 1201 TTGGGAACATAAATGTTCCCGAGGTGATGCTGATTTTCATTTGTCACATGGACTGAA 1260
Db 2465 TTGGGAACATAAATGTTCCCGAGGTGATGCTGATTTTCATTTGTCACATGGACTGAA 2524
QY 1261 TCACATAGATTCTCTCCGTCAGTAACCGTGGGATTAACAATATATGCTTCCAAAGTA 1320
Db 2525 TCACATAGATTCTCTCCGTCAGTAACCGTGGGATTAACAATATATGCTTCC-AAAGTA 2583
QY 1321 TGGACACATCCCAATCAGAAAAAGGTTATCAT 1351
Db 2584 TGGACACATCCCAATCAGAAAAAGGTTATCAT 2614

RESULT 10
ID ABA97932
XX ABA97932 standard; cDNA; 2262 BP.
AC ABA97932;
XX ABA97932;
DT 25-APR-2002 (first entry)
XX Human scavenger receptor-like protein encoding cDNA SEQ ID NO 1.
DE Human; scavenger; receptor; diagnosis; treatment; autoimmune disease;
KW rheumatoid arthritis; gene; ss.
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers

FT CDS 33..2262
FT /*tag= a
FT /product= "scavenger receptor-like protein"
XX
PN JP2001340089-A.
XX
PD 11-DEC-2001.
XX
PF 08-DEC-2000; 2000JP-0375066.
XX
PR 27-MAR-2000; 2000JP-0090772.
XX
PA (SHIO) SHIONOGI & CO LTD.
XX
DR WPI; 2002-144965/19.
DR P-PSDB; ABB08642.
XX
PT New scavenger receptor-like protein for diagnosis, prevention and
PT treatment of autoimmune disease, such as rheumatoid arthritis -
XX
PS Claim 3; Fig 1; 38pp; Japanese.
XX
CC The invention relates to a human scavenger receptor-like protein. The
CC protein is useful as a target molecule for diagnosis, prevention and
CC treatment of autoimmune diseases such as rheumatoid arthritis.
XX
SQ Sequence 2262 BP; 670 A; 576 C; 581 G; 435 T; 0 other;
Query Match 75.9%; Score 1029; DB 24; Length 2262;
Best Local Similarity 100.0%; Pred. No. 1.8e-236;
Matches 1029; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAACAAGATTGATGAGTTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
Db 1234 ATGCAACAAGATTGATGAGTTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 1293
QY 61 ATGGAAGAAATGAAGCTAGTAGCTCCAAAGCATGTGTGCTCATCAAGAAATTTTACAATA 120
Db 1294 ATGGAAGAAATGAAGCTAGTAGCTCCAAAGCATGTGTGCTCATCAAGAAATTTTACAATA 1353
QY 121 CTACAGAGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCAGGAGCCCCCT 180
Db 1354 CTACAGAGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCAGGAGCCCCCT 1413
QY 181 GGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGGGGGAGCCCTGGACACCTTGSCCCT 240
Db 1414 GGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGGGGGAGCCCTGGACACCTTGSCCCT 1473
QY 241 GCGGGTGAGAGAGGCCCAATTTGGACAGCTGGTCCCCCGGAGAGCGTGGCGCAAGGA 300
Db 1474 GCGGGTGAGAGAGGCCCAATTTGGACAGCTGGTCCCCCGGAGAGCGTGGCGCAAGGA 1533
QY 301 TCTAAGGCTCCACAGGGCCCCAAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGGCCCTCAG 360
Db 1534 TCTAAGGCTCCACAGGGCCCCAAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGGCCCTCAG 1593
QY 361 GGCCCCAGTGGGAGCCACAGGGCCCCCGGGCCCAACAGGAGAGGAGTCCCGGCCCTCAG 420
Db 1594 GGCCCCAGTGGGAGCCACAGGGCCCCCGGGCCCAACAGGAGAGGAGTCCCGGCCCTCAG 1653
QY 421 CAGGGCCCTCTGCTGGCTTCAGGGACTTCAGGGCAGCGTTGGGGAGCGCTGGGCTGGA 480
Db 1654 CAGGGCCCTCTGCTGGCTTCAGGGACTTCAGGGCAGCGTTGGGGAGCGCTGGGCTGGA 1713
QY 481 CCTCGGGGACTGCCAGGCTTGGCTGGGGTACCAGGCTGCCAGGCCCCAAGGCCCCCCC 540
Db 1714 CCTCGGGGACTGCCAGGCTTGGCTGGGGTACCAGGCTGCCAGGCCCCAAGGCCCCCCC 1773
QY 541 GGCCCTCTCTGGCCCATCAGGAGCGTGGTGGCCCTTGCCCTGCAGAAATGAGCAACCCCG 600
Db 1774 GGCCCTCTCTGGCCCATCAGGAGCGTGGTGGCCCTTGCCCTGCAGAAATGAGCAACCCCG 1833
QY 601 GCACCGGAGGACAATGGGCTGCCCGCTCACTTGAAGAACTTCACAGACAAATGCTACTAT 660

Db 1834 GCACGGAGGACAAATGGCTGCCCGCTCACTGGAGAACTTCACAGACAAAATGCTACTAT 1893
Qy 661 TTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGCTTTTCTGTGGAAGACAAGTCTTCA 720
Db 1894 TTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGCTTTTCTGTGGAAGACAAGTCTTCA 1953
Qy 721 CATCTTGTTTTCATAAACACTAGAGAGGAGCAAGCAATGGATATAAAAAACAGATGATGGG 780
Db 1954 CATCTTGTTTTCATAAACACTAGAGAGGAGCAAGCAATGGATATAAAAAACAGATGATGGG 2013
Qy 781 AGAGAGAGCCACTGGATCGGCTCACAGACTCACAGCTGAGAGCTGAAATGAATGGAAGTGGTG 840
Db 2014 AGAGAGAGCCACTGGATCGGCTCACAGACTCACAGCTGAGAGCTGAAATGAATGGAAGTGGTG 2073
Qy 841 GATGGACATCTCCAGACTACAAAAATGGAAGCTGGACAGCCGATTAATCGGGTCTAT 900
Db 2074 GATGGACATCTCCAGACTACAAAAATGGAAGCTGGACAGCCGATTAATCGGGTCTAT 2133
Qy 901 GGCCATGGCCAGGAGAAAGACTGTGCTGGTGTGATTTATGCTGGGCACTGGAACGATTC 960
Db 2134 GGCCATGGCCAGGAGAAAGACTGTGCTGGTGTGATTTATGCTGGGCACTGGAACGATTC 2193
Qy 961 CAATGTGAAGACGTCATTAATCTTCAATTTGCGAAAAAGACAGGGACAGTACTGTCTATCT 1020
Db 2194 CAATGTGAAGACGTCATTAATCTTCAATTTGCGAAAAAGACAGGGAGACAGTACTGTCTATCT 2253
Qy 1021 GCATTATAA 1029
Db 2254 GCATTATAA 2262
RESULT 11
AAH43037
ID AAH43037 standard; cDNA; 2637 BP.
XX AC AAH43037;
XX DT 15-OCT-2001 (first entry)
XX DE Nucleotide sequence of a human scavenger receptor.
XX KW Human; scavenger receptor; SRCL-Pl; macrophage; basal immunity;
XX KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
XX KW low density lipoprotein; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 92..2320
XX FT /*tag= a
XX FT /product= "scavenger receptor"
XX PN WO200159107-A1.
XX PD 16-AUG-2001.
XX PF 08-FEB-2001; 2001WO-JP00874.
XX PR 14-FEB-2000; 2000JP-0035155.
XX PR 10-OCT-2000; 2000JP-0309068.
XX PA (FUSO) FUSO PHARM IND LTD.
XX PI Wakamiya N;
XX DR WPI; 2001-497076/54.
XX DR P-PSDB; AAG63347.
XX PT New Scavenger receptor proteins SRCL-Pl with collectin-like structure,
XX PT useful for treatment and diagnosis of diseases associated with oxidized
XX PT low-density lipoprotein accumulation -

PS Claim 6; Page 88-93; 118pp; Japanese.
XX CC The present sequence encodes a human scavenger receptor, designated
CC SRCL-Pl. The SRCL-Pl polypeptide has a collectin-like structure. They
CC are useful in clarifying the functions of macrophages and basal
CC immunity. They are also useful in the treatment, prevention, diagnosis
CC and investigation of diseases such as arteriosclerosis, diabetic
CC complications, bacterial infection and restenosis following angioplasty,
CC which are associated with accumulation of oxidized low density
CC lipoprotein and the binding of advanced glycation end-products into
CC cells.
XX SO Sequence 2637 BP; 829 A; 617 C; 635 G; 556 T; 0 other;
Query Match 66.1%; Score 895.2; DB 22; Length 2637;
Best Local Similarity 81.7%; Pred. No. 9.5e-222;
Matches 1100; Conservative 0; Mismatches 228; Indels 18; Gaps 5;
Qy 1 ATGCAACAAGATTGTGATGAGGTTCGAGTTAGACACTGAAGTAGCCAACTTATCATGTGATT 60
Db 1292 ATGCAGCAAGACATGATGAGGTCAAAGTTAGACACTGAAGTGGCCAACTTATCATGTGTT 1351
Qy 61 ATGGAAGAATAAGTAGTACTCCAAAGCATGCTCAGCTCATCAAGAATTTTACAATA 120
Db 1352 ATGGAAGAAGATGAATACTGGTTGACTCCAAGCACGGTCACTCATCAAGAATTTTACCATT 1411
Qy 121 CTACAAGTCTCCAGCGGCCCGCCAGAGGTCCAAGAGTGCACAGAGGATCCAGGGACCCCT 180
Db 1412 CTACAAGTCTCTGCGCCCGCCAGAGTCCAAAGGTGACAGAGGATCTCAGGGACCACT 1471
Qy 181 GGCCCACTGGCAACAAGGGACAGAAAGAGAGAGAGGGGAGCTTGACCACTGGCCCT 240
Db 1472 GGTCCAACCTGGCAACAAGGGACAGAAAGAGAGAGAGGAGAGCCCTGGTCCACCTGGCCCT 1531
Qy 241 GCGGGTGAGAGAGGCCCAATTTGGACACAGTGTGTCCTCCCGGGAGAGCGTGGCGGCAAGA 300
Db 1532 GCGGGTGAGAGAGGGGCACAAATTTGGACACAGTGTGTCCTCCCGGGAGAGCGTGGCGGCAAGA 1591
Qy 301 TCTAAAGCTCCAGGGGCCCAAGGCTCCGCTGGTTGTCCTGGGAAGCCCGGCCCTCAG 360
Db 1592 TCCAAAGCTCACAGGTTCCCAAGGATCTCGTGGTCCCAAGGAGAGCTGGCCCTCA 1651
Qy 361 GGCCCACTGGGGACCCAGGCCCCCGGCCCAACAGGAGAGAGGAGTCTCCCGGCCCT 420
Db 1652 GGACCTAGTGGGGACCCAGGACCACAGTGTCCACAGCAAGGATGAGACTTCCCTGGCCCT 1711
Qy 421 CAGGCCCTCTGCTTCCAGGGACTTCAGGGCACCTTGGGGACCTTGGGGTGGCTCTGA 480
Db 1712 CAGGCCCTCTGCTTCCAGGGACTTCAGGGCACCTTGGGGTGGAGTACCTCTGA 1771
Qy 481 CTGGGGACTGCCAGGCTTGCCTGGGTACCAAGGATGTCAGGAGCCCAAGGGCCCCCCC 540
Db 1772 CTGGGGGTTGCCAGGCTTGCAGGGGTGTCAGGATGCTTGGGCTTAAGGGACCACT 1831
Qy 541 GGCCCTCTGCGCCATCAGAGCGGTGTGTCCTCCCTGGCCCTGCAGAAATGAGCCAAACCCG 600
Db 1832 GGCCCTTCCAGGCCCTCAGGAGCAATGGAGCCATTGGCTCTGCAGAATGAACCAACCCCA 1891
Qy 601 GCACGGAGGACATGGCTGCCGCTCCTCAGTGAAGAACTTCACAGACAAATGCTACTAT 560
Db 1892 GCATCAGAGGTCAACGGATGTCGCCCTCCTCAGTGAAGAACTTCACAGATAAATGCTACTAT 1951
Qy 661 TTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGCTTTTCTGTGGAAGACAAGTCTTCA 720
Db 1952 TTTTCAGTTGAGAAAGAAATTTTGAGGATGCTAAGCTTTTCTGTGGAAGACAATGCTTCC 2011
Qy 721 CATCTTGTTTTCATAAACACTAGAGAGGAGCAAGCAATGGATATAAAAAACAGATGATGGG 780
Db 2012 CATCTCGTTTTCATAAACACTCAAGAGAGCAAGCAATGGATATAAAAAACAGATGATGGG 2071
Qy 781 AGAGAGAGCCACTGGATCGGCTCACAGACTCACAGCTGAGAGCTGAAATGAATGGAAGTGGTG 840
Db 2072 AGAGAAAGCCATTGGATCGGCTCACAGACTCACAGAAAGAGGAGAAATGGAAGTGGGCTA 2131

Db 772 CTGTGTTTCATTAACACACTAGAGAGGAACAGCAATGGATAAAAAACACAGATGTTAGGGAGA 831

Qy 784 GAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGGAAGTGGCTGGAT 843

Db 832 GAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGGAAGTGGCTGGAT 891

Qy 844 GGGACATCTCAGACTCAAAAAATTGGAAAGCTGGACAGCGGATAAATGGGGTTCATGGC 903

Db 892 GGGACATCTCAGACTCAAAAAATTGGAAAGCTGGACAGCGGATAAATGGGGTTCATGGC 951

Qy 904 CATGGCCAGGAGAGACTGTGGGTTGATTTATGCTGGCGAGTGAACGATTTCCAA 963

Db 952 CATGGCCAGGAGAGACTGTGGGTTGATTTATGCTGGCGAGTGAACGATTTCCAA 1011

Qy 964 TGTGAAGCGTCAATAACTTCATTTGCGAAAAAGACAGGGAGACAGT 1010

Db 1012 TGTGAAGCGTCAATAACTTCATTTGCGAAAAAGACAGGGAGACAGT 1058

RESULT 13

AAH43054

ID AAH43054 standard; DNA; 2256 BP.

XX AC AAH43054;

XX DT 15-OCT-2001 (first entry)

XX DE Nucleotide sequence of a human scavenger receptor.

XX KW Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;

XX KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;

XX KW low density lipoprotein; ss.

XX OS Homo sapiens.

XX FH Key

FT CDS

FT 74..1936

FT /*tag= a

FT /product= "scavenger receptor"

XX PN W0200159107-A1.

XX PD 16-AUG-2001.

XX PF 08-FEB-2001; 2001WO-JP00874.

XX PR 14-FEB-2000; 2000JP-0035155.

XX PR 10-OCT-2000; 2000JP-0309068.

XX PA (FUSO) FUSO PHARM IND LTD.

XX PI Wakamiya N;

XX DR WPI; 2001-497076/54.

XX DR P-PSDB; AAG63350.

XX PT New Scavenger receptor proteins SRCL-P1 with collectin-like structure,

XX PT useful for treatment and diagnosis of diseases associated with oxidized

XX PT low-density lipoprotein accumulation -

XX PS Claim 4; Page 105-109; 118pp; Japanese.

XX CC The present sequence encodes a human scavenger receptor, designated

XX CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They

XX CC are useful in clarifying the functions of macrophages and basal

XX CC immunity. They are also useful in the treatment, prevention, and basal

XX CC and investigation of diseases such as arteriosclerosis, diabetic

XX CC complications, bacterial infection and restenosis following angioplasty,

XX CC which are associated with accumulation of oxidized low density

XX CC lipoprotein and the binding of advanced glycation end-products into

XX CC cells.

XX SQ Sequence 2256 BP; 721 A; 524 C; 528 G; 483 T; 0 other;

Query Match 54.9%; Score 743.8; DB 22; Length 2256;
Best Local Similarity 91.9%; Pred. No. 1.6e-182;
Matches 797; Conservative 0; Mismatches 67; Indels 3; Gaps - 1;

Qy 489 ACTGCCAGGCTTGGCTGGGTACAGGCTACAGGCGCCCAAGAGGCCCGCCCGGCCCTCC 548

Db 1393 ACTACAAGTGTCCACGGGCCCCAGGGTCCAAGAGGTGACAGAGGATCCAGGGACCC 1452

Qy 549 TGGCCCATCAGAGGGGTGGCCCTGGCCCTGCAGAAATGAGCCAAACCCCGCACCGGA 608

Db 1453 TGGCCCAACTGCAACAAGGGACAGAAAGAGAGAGGGGGAGCGCTGGAC---CACCTGG 1509

Qy 609 GGACAAATGGCTGCCCGCTCTACTGGAAGAACTTTCACAGACAAATCTCTACTATTTTCA 668

Db 1510 CCTCGGGCTGCCCGCTCTACTGGAAGAACTTTCACAGACAAATCTCTACTATTTTCA 1569

Qy 669 TGAGAAAGAAATTTTGGAGGATGCAAGCTTTCTGTGTGAAGCAAGCTTTCACATCTTGT 728

Db 1570 TGAGAAAGAAATTTTGGAGGATGCAAGCTTTCTGTGTGAAGCAAGCTTTCACATCTTGT 1629

Qy 729 TTTCAATAACACTAGAGAGAACAGCAATGGATAAAAAACAGATGTTAGGGAGAGAG 788

Db 1630 TTTCAATAACACTAGAGAGAACAGCAATGGATAAAAAACAGATGTTAGGGAGAGAG 1689

Qy 789 CCAGTGGATCGGCTTCACAGACTCAGAGCGTGAATGAATGGAATGGCTGGATGGGAC 848

Db 1690 CCAGTGGATCGGCTTCACAGACTCAGAGCGTGAATGAATGGAGAGTGGCTGGATGGGAC 1749

Qy 849 ATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAATGGCCATGG 908

Db 1750 ATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAATGGCCATGG 1809

Qy 909 GCCAGGAGAACTGTGCTGGGTTGATTTATGCTGGGAGTGGAAACGATTTCCAATTTGA 968

Db 1810 GCCAGGAGAACTGTGCTGGGTTGATTTATGCTGGGAGTGGAAACGATTTCCAATTTGA 1869

Qy 969 AGAGCTCAATAACTTCATTTGCGAAAAAGACAGAGGAGACAGTACTGTCTGTCATTATA 1028

Db 1870 AGAGCTCAATAACTTCATTTGCGAAAAAGACAGAGGAGACAGTACTGTCTGTCATTATA 1929

Qy 1029 ACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGTCTCTCAAAGGCAAGGACACTCCT 1088

Db 1930 ACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGTCTCTCAAAGGCAAGGACACTCCT 1989

Qy 1089 TTCTAATTCATCACCTTCTCATCAGATTGAAAAAAGACACTGAAAAACCAATTAC 1148

Db 1990 TTCTAATTCATCACCTTCTCATCAGATTGAAAAAAGACACTGAAAAACCAATTAC 2049

Qy 1149 TGAATAAATTTGACAGCTAGTGTGTTTATCCATCCGTCATTACCAAGACTTTGGGAAC 1208

Db 2050 TGAATAAATTTGACAGCTAGTGTGTTTATCCATCCGTCATTACCAAGACTTTGGGAAC 2109

Qy 1209 TAAATGTTCCCGAGGTGATGCTGATTTTCATTTGTCACATGGATGATACATACATAG 1268

Db 2110 TAAATGTTCCCGAGGTGATGCTGATTTTCATTTGTCACATGGATGATACATACATAG 2169

Qy 1269 ATTCTCCTCCCTCAGTACCGGTGATTTATCAAAATTTATGTTCCAAAGTATGAAACAC 1328

Db 2170 ATTCTCCTCCCTCAGTACCGGTGATTTATCAAAATTTATGTTCCAAAGTATGAAACAC 2229

Qy 1329 TCCAATCAGAAAAAGTTTATCATCCG 1355

Db 2230 TCCAATCAGAAAAAGTTTATCATCCG 2256

RESULT 14

AA571131/c

ID AA571131 standard; cDNA; 493 BP.

XX AC AA571131;

XX DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #6935.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS WO200175067-A2.
PN 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX P-PSDB; ABG06944.
DR New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 1; SEQ ID No 6935; 103pp; English.
PS The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 493 BP; 121 A; 116 C; 99 G; 157 T; 0 other;
SQ Query Match 30.5%; Score 413; DB 23; Length 493;
Best Local Similarity 97.8%; Pred. No. 4.7e-97;
Matches 482; Conservative 0; Mismatches 5; Indels 6; Gaps 6;
QY 812 CAGACGGTGAATAAGTGAAGTGGCTGGATGGGACATC-TCCAGACTACA-AAAATTG 869
DB 493 CAGACGGTGAATAAGTGAAGTGGCTGGATGGGACATC-TCCAGACTACA-AAAATTG 434
QY 870 GAAAGCTGGAC-AGCGGATAACT-GGGTCATGCCATGGCCAGGAGAGACTGTGCT 927
DB 433 GAAAGCTGGCAAGCCGAGTAACTACTGGGGTCAATGGCCAGGAGAGACTGTGCT 374
QY 928 GGGTTGATTATGCTGGCAGTGGACGATTTCCAAATGTGAAGACGTCATTAACATTCATT 987
DB 373 GGGTTGATTATGCTGGCAGTGGACGATTTCCAAATGTGAAGACGTCATTAACATTCATT 314
QY 988 TCCGAAAAGACAGGAGACAGTACTGTCATTCATTAACGAGCTGTGATGGATCA 1047
|||||

Db 313 TCGAAAAAGACAGGAGACAGTACTGTCTATCTGCATTATATAACGAGCTGTGATGGATCA 254
QY 1048 CATGAGCAAAATTTTCAGCTCTCAAAGGCAAGGACACTCTCTTAATTGATACCTTC 1107
|||||
Db 253 CATGAGCAAAATTTTCAGCTCTCAAAGGCAAGGACACTCTCTTAATTGATACCTTC 194
QY 1108 TCATCAGATTGAAAAAAGACACTGAAACCAATTAATGAAAAAATTTGACAGCT 1167
|||||
Db 193 TCATCAGATTGAAAAAAGACACTGAAACCAATTAATGAAAAAATTTGACAGCT 134
QY 1168 AGTGTGTTTACCATCCGTCATTACCCAAAAGACTTTGGGAACATAAATGTTCCCA 1226
|||||
Db 133 AGTGTGTTTACCATCCGTCATTACCCAAAAGACTTTGGGAACATAAATGTTCCCA 74
QY 1227 GATATGCT-GATTTTCATTGTGCACATGACGACGAATCACAATGATTCCTCCGTCAGTA 1285
|||||
Db 73 GATATGCTGGATTTCATTGTGCACAGGGACTGAATCACAATGATTCCTCCGTCAGTA 14
QY 1286 ACCGTGCGATTAT 1298
|||||
Db 13 CCGGTGCGATTAT 1
RESULT 15
ABAT17153
ID ABAT17153 standard; DNA; 5041 BP.
XX AC ABAT17153;
XX DT 23-JAN-2002 (first entry)
XX DE Human nervous system related polynucleotide SEQ ID NO 9484.
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antistickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiparasitic; cardiatic; antidiabetic; antilulcer; anticonvulsant; antifungal;
KW neurological disease; infection; immune disorder; cardiovascular disorder;
XX neurologic disease; infection; nephrotropic; gene therapy; vaccine; ds.
OS Homo sapiens.
XX WO200159063-A2.
XX PD 16-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01334.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0232081.
 PR 14-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 20-OCT-2000; 2000US-0242221.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-541565/60.
 XX
 DR Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 XX useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 PT
 XX
 PS Disclosure: SEQ ID NO 9484; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (AB11004-ABA21534) and proteins
 CC (AB114678-AB118001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 5041 BP; 1480 A; 1074 C; 1103 G; 1384 T; 0 other;

[illegible]

Search completed: March 21, 2003, 02:19:58
Job time : 300.408 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 06:54:47 ; Search time 3616.46 Seconds
(without alignments)
4401.885 Million cell updates/sec

Title: US-09-763-712A-2

Perfect score: 2969

Sequence: 1 MYSHNVIMNINLNLTVQV.....EDVNNFICEKDRETVLSSAL 547

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-O=/cgn2_1/usptc_pool/US09763712/runat_14032003_100949_18101/app_query.fasta_1.1877
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cd1 -LIST=45
-DOCLALIGN=200 -THR_SCORE=pct -THR_WAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712@cgn_1_1.6828_runat_14032003_100949_18101 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_other:*
33: em_hgt_mus:*
34: em_hgt_pln:*
35: em_hgt_rod:*
36: em_hgt_mam:*
37: em_hgt_vrt:*
38: em_sy:*
39: em_hgtgo_hum:*
40: em_hgtgo_mus:*
41: em_hgtgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|--------|---------------|--------|-------|-----------------------------|
| 1 | 2369 | 100.0 | 2983 | 9 | AB005145 Homo sapi |
| 2 | 2363 | 99.8 | 2641 | 6 | AX047353 Sequence |
| 3 | 2363 | 99.8 | 3058 | 9 | AB038518 Homo sapi |
| 4 | 2765 | 93.1 | 2637 | 10 | AB078434 Mus muscu |
| 5 | 2751 | 92.7 | 3291 | 10 | AB038519 Mus muscu |
| 6 | 2693 | 90.7 | 4330 | 9 | AB052103 Homo sapi |
| 7 | 2688 | 90.5 | 2005 | 6 | AX544442 Sequence |
| 8 | 2688 | 90.5 | 2005 | 6 | AX490920 Sequence |
| 9 | 2309 | 77.8 | 1886 | 9 | HSM802985 |
| c 10 | 1239 | 41.7 | 169088 | 2 | AC016128 Homo sapi |
| 11 | 1239 | 41.7 | 178022 | 2 | AP001022 Homo sapi |
| 12 | 1239 | 41.7 | 187635 | 2 | AP001022 Homo sapi |
| c 13 | 1239 | 41.7 | 188439 | 2 | AP000935 Homo sapi |
| 14 | 1212 | 40.8 | 188255 | 2 | AP000939 Homo sapi |
| c 15 | 1168 | 39.3 | 182029 | 2 | AC114677 Mus muscu |
| c 16 | 1168 | 39.3 | 193208 | 2 | AC102618 Mus muscu |
| 17 | 1021.5 | 34.4 | 130763 | 2 | AC112416 Rattus no |
| 18 | 991 | 33.4 | 71044 | 2 | AC024368 Homo sapi |
| 19 | 981 | 33.0 | 182029 | 2 | AC114677 Homo sapi |
| 20 | 981 | 33.0 | 193208 | 2 | AC102618 Mus muscu |
| c 21 | 842 | 28.4 | 188255 | 2 | AP000939 Homo sapi |
| 22 | 765 | 25.8 | 3636 | 9 | AB007829 Homo sapi |
| 23 | 765 | 25.8 | 3685 | 6 | E32511 Scavenger r |
| 24 | 765 | 25.8 | 3810 | 6 | E32509 Scavenger r |
| 25 | 671 | 22.6 | 2215 | 10 | BC026446 Mus muscu |
| c 26 | 621 | 20.9 | 188637 | 2 | AL845306 Danio rer |
| 27 | 491.5 | 16.6 | 1265 | 10 | RATSPD M81231 Rat pulmona |
| 28 | 483.5 | 16.3 | 3508 | 3 | AF053538 Alvinella |
| 29 | 481.5 | 16.2 | 1183 | 10 | BC003705 Mus muscu |
| 30 | 481.5 | 16.2 | 1253 | 10 | MUSSPD L40156 Mus musculu |
| 31 | 474.5 | 16.0 | 1410 | 6 | AX334792 Sequence |
| 32 | 474.5 | 16.0 | 1410 | 6 | HSNRNAPD X65018 H.sapiens m |
| 33 | 474 | 16.0 | 1454 | 4 | BTLSPD X75911 B.taurus mR |
| 34 | 473.5 | 15.9 | 1284 | 4 | AF509590 Bos tauru |
| 35 | 473.5 | 15.9 | 1301 | 9 | BC022318 Homo sapi |
| 36 | 471 | 15.9 | 2377 | 9 | BC008760 Homo sapi |
| 37 | 471 | 15.9 | 5676 | 6 | E07265 E07265 cDNA encodi |
| 38 | 471 | 15.9 | 5676 | 6 | HUMCALV D90279 Human mRNa |
| 39 | 471 | 15.9 | 7138 | 9 | HUMPAIV M76729 Human pro-a |
| 40 | 465 | 15.7 | 6114 | 10 | CRUPAIV M76730 Chinese ham |
| 41 | 462.5 | 15.6 | 4428 | 6 | AX146422 Sequence |
| 42 | 462.5 | 15.6 | 4428 | 6 | AX146424 Sequence |
| 43 | 462.5 | 15.6 | 4990 | 10 | CRUCOLVITA L06863 Cricetus |
| 44 | 462 | 15.6 | 2777 | 5 | GU07973 Gallus gall |
| 45 | 462 | 15.6 | 5575 | 5 | AF137273 Gallus ga |

ALIGNMENTS

... RESULT 1

AB005145
LOCUS AB005145 2983 bp mRNA linear PRI 21-NOV-2001
DEFINITION Homo sapiens CL-PI mRNA for collectin placenta 1, complete cds.
ACCESSION AB005145
VERSION AB005145.1 GI:17026100
KEYWORDS
SOURCE Homo sapiens female tissue_lib:placenta cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H.,
Sakai,Y., Fukuch,A., Sakamoto,T., Itabe,H., Suzutani,T.,
Ogasawara,M., Yoshida,I. and Wakamiya,N.
The membrane-type collectin CL-PI is a scavenger receptor on
vascular endothelial cells
J. Biol. Chem. 276 (47), 44222-44228 (2001)
21570232
REFERENCE 2 (bases 1 to 2983)
AUTHORS Ohtani,K.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1997) Katsuki Ohtani, Asahikawa Medical College,
Department of Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa,
Hokkaido 078-8510, Japan (E-mail: ohtani@asahikawa-med.ac.jp,
Tel:+81-166-68-2393, Fax:+81-166-68-2399)
FEATURES
source
gene
CDS
1..2983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="female"
/tissue_lib="placenta"
1..2983
/gene="CL-PI"
71..2299
/gene="CL-PI"
/codon_start=1
/product="collectin placenta 1"
/protein_id="BAB72147.1"
/db_xref="GI:17026101"
/translation="MKDDFAEEBVEQSVFGYKREGIOEGTQCKKNWALKFISILLY
ILCATITVAIGYKVKMDNVTGSMETSRQYDDKLTAVESDLKILGDDQNGKAI
SYNSELSTSPRSIDLDLROOLRETXETSKNKLLEKQASGDALVDQSOLETLENN
SELTITVNTKLYANGYVYVNLQDQTSVLQNLQNMYSHNVIWNLNLNLTVQNSLA
LIITNLQRSDDTSQA LQRIKDNFQNLQQVFLQAKDQTDMLKEKVQSLQTLAANSALA
KANNDTVDMSQNSLSTFQMENTITTSQANEQLKDLQDLHKVDANRTAIKFNQLEE
RFQLFTDVIINISYTAHILRTLSNLNEVTTCTDTLTAKHTDFTLSLNLNTANI
RLDSVLMOODLMRSLRTEAVANLSVIMEEMKLVDSKHGLIKNFTILQGPGRGP
RGDRSGKPGPTGNGKQGEKGPAGPGRGPAGPAGPGRGKSGKSGQPK
RGDRSGKPGQSGDGPQPGKGLPGQPGPGFQGLQCTGVEGPGVPGRGLPG
LPGVGMGPKGPPGPGSGAVVPLAQNLQEPAPEDNGCGPHWKNFTDKCYFFVE
KEIFDAKFLCEDKSHLVPINTREBQWIKQMVGRSHWIGLTSERENENKWLDDG
TSPDYKNWKAQPDNNGHGGEDCAGLIYAGQWDFQEDVYNNFICEKDRETVLSG
AL"

BASE COUNT 914 a 707 c 703 g 659 t
ORIGIN

Alignment Scores:
Pred. No.: 1.8e-95 Length: 2983
Score: 2969.00 Matches: 547
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-763-712A-2 (1-547) x AB005145 (1-2983)

Qy 1 MetTyrSerHisAsnValIleMetAsnLeuAsnLeuLeuThrGlnValGln 20
|||||
Db 656 ATGTATTCTATAATGTGGTCATCATGACCTCAACAACTGAACCTGCCAGGTCAG 715
|||||

Qy 21 GlnArgAsnLeuLeuThrAsnLeuGlnArgSerValAspAspThrSerGlnAlaIleGln 40
|||||

Db 716 CACAGGAACCTCATCAGAACTGCACGGCTCTGTGATGATGACACAAAGCCAGGCTATCCAG 775
|||||


```
Qy 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpPlyAsnPheThr 420
|||||
Db 1856 AATGAGCCAAACCCGGCACCAGGACAAATGGCTGCCCGCCTCACTGGAAGAACTTCACA 1915
|||||
Qy 421 AspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
|||||
Db 1916 GACAAATGCTACTATTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAAAGCTTTTCTGT 1975
|||||
Qy 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
|||||
Db 1976 GAAGACAAGTCTTCACATCTGTTTTCATAAACACTAGAGAGAACACCAATGGATAAAA 2035
|||||
Qy 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
|||||
Db 2036 AAACAGATGGTAGGAGAGAGACCACTGGATCGGCCTTCACAGACTCAGAGCGTGAANAAT 2095
|||||
Qy 481 GluTrpLysTyrLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyClnPro 500
|||||
Db 2096 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCGC 2155
|||||
Qy 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
|||||
Db 2156 GATAACTGGGTCTATGGCCATGGCCAGGAGAGACTGTGCTGGTTGATTTATGCTGG 2215
|||||
Qy 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
|||||
Db 2216 CAGTGGAAACGATTTCCAAATGTGAAGACGTCAATAAATCACTTTCATTTGCGAAAAAGACAGGGAG 2275
|||||
Qy 541 ThrValLeuSerSerAlaLeu 547
|||||
Db 2276 ACAGTACTGTCTATCTGCATTA 2296
|||||

RESULT 2
AX047353
LOCUS AX047353 2641 bp DNA linear PAT 15-DEC-2000
DEFINITION Sequence 39 from Patent WO0068380.
ACCESSION AX047353
VERSION AX047353.1 GI:11876599
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2641)
AUTHORS Bandman,O., Hillman,J.L., Tang,Y.T., Lal,P., Yue,H., Baughn,M.R.,
Lu,D.A. and Azimzai,Y.
TITLE Extracellular matrix and adhesion-associated proteins
JOURNAL Patent: WO 0068380-A 39 16-NOV-2000;
KEYWORDS Incyte Genomics, Inc. (US)
FEATURES
source
1..2641
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 4586187CBI"

BASE COUNT 823 a 628 c 608 g 582 t
ORIGIN

Alignment Scores:
Pred No.: 2,64e-95 Length: 2641
Score: 2963.00 Matches: 546
Percent Similarity: 99.82% Conservative: 0
Best Local Similarity: 99.82% Mismatches: 1
Query Match: 99.80% Indels: 0
DB: 6 Gaps: 0

US-09-763-712A-2 (1-547) x AX047353 (1-2641)

Qy 1 MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
|||||
Db 297 ATGTATTCTCATATGTGGTTCATCATGAACCTCAACACACCTGAACCTGACCCAGCGTGAG 356
|||||
Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
|||||
```

```
Db 357 CAGAGGAACCTCATCAGCAATCTGCAGCGGTCTGTGGATGACACAAAGCAGGCTATCCAG 416
|||||
Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
|||||
Db 417 CGAATCAGAAGACGACTTTCAAAATCTGCAGCAGGTTTTTCTCAAGCCCAAGAGGACAGC 476
|||||
Qy 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80
|||||
Db 477 GATTGGCTGAAGGAGAAAGTGCAGAGCTTGACAGCGCTGGCTGCCAACAACCTCTCGCTTG 536
|||||
Qy 81 AlaIysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
|||||
Db 537 GCCAAAGCCACAACAGCACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGT 596
|||||
Qy 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
|||||
Db 597 CAGATGGAGACATCACCACATCTCTCAAGCCACAGGAGAGAACTGAAAGACCTGGCAG 656
|||||
Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
|||||
Db 657 GACTTACACAAGATGCAGAGAAATAGAACAGCATCAAGTTCAACCAACTGGAGGAACGC 716
|||||
Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
|||||
Db 717 TTCCAGCTCTTTGAGACCGGATATGTGAACATCATTTAGCAATATCAGTTACACAGCCCAC 776
|||||
Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValAlaArgThrThrCysThrAspThrLeu 180
|||||
Db 777 CACCTGCGGACGCTGACCAGCAATCTAATGAAGTCAGGACCACCTTGACAGATACCCTT 836
|||||
Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
|||||
Db 837 ACCAAACACACAGATGATCTGACCTCTTGAATAATAATCCCTGGCCCAACATCCGTTTGGAT 896
|||||
Qy 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
|||||
Db 897 TCTGTTCTCTCAGGATGCAACAGATTTGATGAGTTCGAGGTTAGACACTGAAGTAGCC 956
|||||
Qy 221 AsnLeuSerValIleMetGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
|||||
Db 957 AACTTATCAGTGATTATGGAAGAAATGAAGTAGTAGACTCCACAGCATGGTCAGCTCATC 1016
|||||
Qy 241 LysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGly 260
|||||
Db 1017 AAGAAATTTTACAATACTACAAGTCCACCGGCCCCAGGGGTCCTCAAGAGGTGCACAGAGA 1076
|||||
Qy 261 SerGlnGlyProProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro 280
|||||
Db 1077 TCCCAGGAGCCCCCTGGGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGAGGGAGCGCT 1136
|||||
Qy 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlu 300
|||||
Db 1137 GGACCACCTGGCCCTGCGGGTGAGAGAGGCCCAATTTGACACAGCTGGTCCCCCGGAGAG 1196
|||||
Qy 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
|||||
Db 1197 CGTGGCGGCAAAAGGATCTAAAGGCTCCAGGGGCCCAAAAGGCTCCCGTGGTCCCTGGG 1256
|||||
Qy 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
|||||
Db 1257 AAGCCCGCCCTTCAGGGGCCCAAGTGGGACCCAGGCCCGCCCGGCCCAAGGAGAGAG 1316
|||||
Qy 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
|||||
Db 1317 GGACTCCCCGGCCCTCAGGGCCCTCTGGCTTCCAGGGACTTCAGGGCACCCGTTGGGGAG 1376
|||||
Qy 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
|||||
Db 1377 CCTGGGGTGCCTGGACCTCGGGGACTGCAGGCTTTGCCTGGGGTACCAGGCATGCCAGC 1436
|||||
Qy 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
|||||
Db 1437 CCCAAGGGCCCCCGGCCCTCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAG 1496
|||||
```

QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
|||||
Db 1497 AATGAGCAACCCGCGCAGGAGCAATAGTCTGCCCGCTCACTGGGAAGACTTCACA 1556
QY 421 AspLysCysTyrrTyPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
|||||
Db 1557 GACAATGCTACTATTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCTTTCTCT 1616
QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
|||||
Db 1617 GAAGACAAGCTCTTCATCTTTTCATAAACACTAGAGAGAGACAGCAATGGATAAAA 1676
QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
|||||
Db 1677 AAACAGATGCTAGGAGAGAGAGCCACTGATCGGCTCACAGACTCAGAGCGTGAATAAT 1736
QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrrLysAsnTrpLysAlaGlyGlnPro 500
|||||
Db 1737 GAATGGAAGTGGCTGATGGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCG 1796
QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrrAlaGly 520
|||||
Db 1797 GATACTGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGTTGATTTATGCTGGG 1856
QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
|||||
Db 1857 CAGTGGACGATTTCCATGTGAAGCGTCAATACTTCATTTGCCAAAAAGACAGGGAG 1916
QY 541 ThrValLeuSerSerAlaLeu 547
|||||
Db 1917 ACAGTACTGTCTATCTGCATTA 1937
RESULT 3
AB038518
LOCUS
DEFINITION Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
type 1, complete cds.
ACCESSION AB038518
VERSION AB038518.1 GI:13365514
KEYWORDS
SOURCE
ORGANISM Homo sapiens tissue_lib:Placenta cdna to mRNA.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
Nakamura, K., Funakoshi, H., Miyamoto, K., Tokunaga, F. and Nakamura, T.
Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRCL), a novel member of a
scavenger receptor family
Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
21092718
2 (bases 1 to 3058)
Nakamura, K. and Nakamura, T.
Direct Submission
Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical
School, Division of Biochemistry, Biomedical Research Center; 2-2
Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail: knakamura@bich.med.osaka-u.ac.jp,
Tel: 81-6-6879-3783 (ex.3783), Fax: 81-6-6879-3789)
Location/Qualifiers
1. .3058
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_lib="Placenta"
1. .3058
/gene="SRCL"
209. .2437
/codon_start=1
/product="scavenger receptor with C-type lectin type I"
/protein_id="BAB39147.1"
/db_xref="GI:13365515"

/translation="MKDFAEEVEVQSEYKRFQIQEQTCTCKKNNWALKPSIILLY
ILCALLITVAILGYKVVKMDNVTGKMETSRQTYDDKLITAVESDKLGLDQGTGKAI
STNSLSTFRSDILDROOLREITEKTKNDKTLKLEKQSDALVDROSLQKLENN
SFLITVNTKLTQAVNGYVTNLQODTSVLQGNLONOMYSHNVVIMNLLNLTOVOORN
LITNLQPSVDDTSQAIOIKNDRONLOQFLQAKKDTMLKEKQVSLQTLAANNSALA
KANNDLEDMSQNSFTGQMENITTSQANEONLKDLDLHKDAENRTAKFNQLEE
RFQIFETDIVNIIISNISTAHRLTLTSLNEVKTCTDTLTTKHTDDTLTSLNNTLANI
RLDSVSLRMOQDLMSRLDTEVANLSVMEEMKLVDSKHQGLIKNFTLLOQPPGPRGP
GSRGSGQPPGPTNGKQGEKGPPEGPAGERPPIGAPGPGERGKSGKSGQPK
LPGVPMGPKPGPPGPGSCAVVPLALQNSPTPAPEDNSCPHPHWKFTDKCYFSVE
KEIFEDAKLFCEDKSSHLYFINTREBQQWIKQWVRESHWIGLITDSRENEWKWDG
TSPDYKNWKAGQPDNWHGHGPGDEDCAGLIYAGQWDFQCEDVNNFICEKORETVLSS
AL"
polyA_signal 3020. .3025
/gene="SRCL"
BASE COUNT 911 a 744 c 764 g 639 t
ORIGIN
Alignment Scores:
Pred. No.: 2,97e-95 Length: 3058
Score: 2963.00 Matches: 546
Percent Similarity: 99.82% Conservative: 0
Best Local Similarity: 99.82% Mismatches: 1
Query Match: 99.80% Indels: 0
DB: 9 Gaps: 0
US-09-763-712A-2 (1-547) x AB038518 (1-3058)
QY 1 MetTyrrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
|||||
Db 794 ATGTATTCATATAATGTGGTCTATCATCAACCTCAACAACCTGAACCTGACCCAGGTGCAG 853
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspAspThrSerGlnAlaIleGln 40
|||||
Db 854 CAGAGGAACCTCATCAGCAATCTGCAGCGTCTGTGGATGACACAGCCAGGCTATCCAG 913
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
|||||
Db 914 CGAATCAAGACGACTTTCAAAATCTGCAGCAGGTTTTCTTCAAGCCCAAGAGGACAGC 973
QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
|||||
Db 974 GATTGGCTGAAGAGAGAAAGTGCAGAGCTTGCAGACACTGGCTGCCAACACTCTGCGTTG 1033
QY 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
|||||
Db 1034 GCCAAGCCCAACACGACACACCTGGAGGATATGAACAGCCAGCTCAACTCATTTCAGGT 1093
QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
|||||
Db 1094 CAGATGGAGAACATCACCACTATCTCTCAAGCCCAACGAGCAGACCTGAAAGACCTGCAG 1153
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
|||||
Db 1154 GACTTACACAAAGATGCGAGAGATAGACAGCCATCAAGTTCAACCACTGGAGGAACCC 1213
QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrrThrAlaHis 160
|||||
Db 1214 TTCCAGCTCTTTGAGACGGATATTGTGAACATCATTAGCAATATCAGTTACACAGCCAC 1273
QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
|||||
Db 1274 CACCTGGCGAGCGCTGACCAAGCAATCTAAATGAAGTCAGGACCACCTTGACACAGATACCCCT 1333
QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
|||||
Db 1334 ACCAAACACACAGATGATCTGACCTCTTGAATAATACCTGGCCCAACATCCGTTTGGAT 1393
QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
|||||
Db 1394 TCTGTTTCTCTCAGGATGCCAACAGANTTTGATGAGGTGAGGTTAGACACTGAAGTAGCC 1453

AB078434.1 GI:21901968

Mus musculus tissue_lib:Liver cdna to mRNA.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathia; Muridae; Murinae; Mus.

1 Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Keshi, H., Sakai, Y., Fukuoh, A., Sakamoto, T., Itabe, H., Suzutani, T., Ogasawara, M., Yoshida, I. and Wakamiya, N.

cDNA cloning of mouse CL-P1 gene

Unpublished

2 (bases 1 to 2637)

Ohtani, K., Fukuoh, A., Eda, S., Kawai, T., Kase, T., Keshi, H., Sakai, Y., Fukuoh, A., Sakamoto, T. and Wakamiya, N.

Direct Submission

Submitted (18-JAN-2002) Katsuki Ohtani, Asahikawa Medical College, Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa, Hokkaido 078-8510, Japan (E-mail: ohtani@asahikawa-med.ac.jp, Tel:81-166-68-2393, Fax:81-166-68-2399)

FEATURES

Location/Qualifiers

1..2637

/organism="Mus musculus"

/db_xref="taxon:10090"

/tissue_lib="Liver"

1..2637

/gene="CL-P1"

92..2320

/gene="CL-P1"

/codon_start=1

/product="collectin placenta 1"

/protein_id="BAC05523.1"

/db_xref="GI:21901969"

/translation="MKDDFAEEVQSFYKRGIQEGTQCTCKCKNNWALKFSLVILYILCALPITVAIIIGYKVERKMDNVDMGEMTSHTYIDNKLFAVEDSLKKGQAGKKAALSTNSELSTFSDILDURQQLQETETKSKNDLEKLAQNGSDLSVRQSGLKETLNALSFLLTVNTKQLAYNGVTYVNLQDTSVLGNLQSQVSVIMNLNLNLTOVQRNLSKNNLOQSDVDSIARIQNFQNLQVFLQAKDFTDLKDKVQSLQTLAANNSALAICANLQDMDNSLSPFGMDNITTSQANEOSLKDLDLHKDTEKRTAVKFSQLEERFOVFETDIVNIIISNYSYTAHLTLTSLNDVYRTCTDPTLTHTDLTSLNNTLVNIRFVLSLRMQQDMRSLKQDEKGEPPGPPGAGERGTIGVPGPGEKSGKSGSGKPGKDRSGSPGKPGQSGDPGPPGPGKDLPGQPGFLOQTGVGPGVGPGRPLGLPVGMPGPGSGPPGPGSGAMEPLALQNEPTPASEVNGCPPHKWNFTDKCYFSLKEITFDEAKLFCEDKSHLVFINSRSEQQWIKHTVGRSHWIGLTDSEQSEKWLQDSPEVDYKNWKAQPDNMGSGHGPGECDAGLIYAGQWDFQCEINFFICEREREAVPSSIL"

BASE COUNT 829 a 617 c 635 g 556 t

ORIGIN

Alignment Scores:

Pred. No.: 1,97e-88 Length: 2637

Score: 2765.00 Matches: 502

Percent Similarity: 96.71% Conservative: 27

Best Local Similarity: 91.77% Mismatches: 18

Query Match: 93.13% Indels: 0

DB: 10 Gaps: 0

US-09-763-712A-2 (1-547) x AB078434 (1-2637)

Qy 1 MetTyrSerHisAsnValIleMetAsnLeuAsnLeuAsnLeuThrGlnValGln 20

||||||| :|||

Db 677 ATGTATTCTCAGACGGTGGTTATCATGAACCTCAACACCTGAACCTAACCGGTTCCAG 736

Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40

||||||| :|||

Db 737 CAGAGGACACCTTATCTCAAAATCTCAGCAGGCTCTGTGGATGACACAAAGCGCTGGCCATCCAG 796

Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60

||||||| :|||

Db 797 CGAATTAAAGATGATTTCCTCAAAATCTGCAGCAGGTTTTCTTCAAGCCAAAGAGACACC 856

Qy 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80

QY 41 ArgTleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
|||||
797 CGAATTAAAGATGATTCCAAATCTGCAGCAGGTTTTCTTCAAGCCAAAGAGCAC 856
DB
QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaIaLaAsnAsnSerAlaLeu 80

```
Db 857 GATTGGCTAAAGAAAGTACAGAGCTTGCAGACATTGGCTGCCAACAACTCTGCCCTG 916
QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
Db 917 GCCAAAGCCAAACATCACACCCTAGAGATATGAATAGCAGCTCAGCTCATTACAGGT 976
QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
Db 977 CAGATGGACAACATTACCACTATCTCACAGGCCAACGAGCAGAGCCCTGAAGACCTTCAG 1036
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
Db 1037 GACTTACACAGGATACAGAAATAGACAGCTGTCAAGTTGAGCCAACTTGAGGAAGCC 1096
QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerIleThrAlaHis 160
Db 1097 TTCCAGGCTTTGAGACACATATTGTGAACATCATTTAGCAACATCAGCTACAGCCCAT 1156
QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
Db 1157 CACCTGAGGACACTGACCAAGCAATCTGAATGATGTTAGGACCACATGCACAGACACCTTG 1216
QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrIleAlaAsnIleArgLeuAsp 200
Db 1217 ACCAGACACAGGATGACCTGACCTCCTTGAATAACACACTAGTCAACATCCGCTTGGAT 1276
QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
Db 1277 TCTATTCTCTCAGGATGACGACAGACATGATGAGTCAAAAGTTAGACACTGAAGTGGCC 1336
QY 221 AsnLeuSerValIleMetGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
Db 1337 AACTTATCAGTGGTATGAAGAGATGAACCTGTTGACTCAAGCACGCTCAGCTCATC 1396
QY 241 LysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGly 260
Db 1397 AAGAACTTTACCATTTACAAAGTCTCTGCGCCCGAGAGGTCCAAAGAGTGCACAGAGA 1456
QY 261 SerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGlyLysGlyGluPro 280
Db 1457 TCTCAGGACCACTGGTCCAACTGCGACACAGGGACAGAGGAGAGAGAGCT 1516
QY 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlu 300
Db 1517 GTTCCACCTGGCCCTCGCGGTGAGAGGGGCACAAATTGGACCACTCGCGCCCTCTGAGAG 1576
QY 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
Db 1577 CGTGGCAGCAAAAGGATCCAAAGGCTCACAGGTTCCCAAGGATCTCGTGGGTCCTCCAGGG 1636
QY 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
Db 1637 AAGCCTGGCCCTCAGGACCTAGTGGGAGCCACAGGACCACTCCAGGCAAGGAT 1696
QY 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
Db 1697 GGACTCCCTGGCCCTCAGGCGCTCTCTGGCTTCCAGGGACTACAGGCACTCTGGGTGAG 1756
QY 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
Db 1757 CTTGGAGTACCTGGAGCTCGGGGTTGCCAGGCTTGGCCAGGGGTGCCAGGCATGCTGGG 1816
QY 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
Db 1817 CCTAAGGACCACTGGCCCTCCAGGCCCTCAGGACCAATGGAGCCATGGCTCTGCGAG 1876
QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
Db 1877 AATGAACCAACCCAGCATCAGAGGTCAACGGATGTGCCGCTCAGTGGGAAGAACTTCACA 1936
QY 421 AspLysCysTyTyPheSerValGluLysGluIlePheGluAlaLysLeuPheCys 440
```

```
Db 1937 GATAAATGCTACTATTTTTCATTGGAAAAAGAAATTTTGAAGATGCTAAGCTTTTCTGT 1996
QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
Db 1997 GAAGACAAATCTCCCATCTCGTTTCATAAACTCAAGAGAAGAACAGCAATGGATAAAA 2056
QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
Db 2057 AAGCATACCGTGGGAGAGAAAGCCATTGTCATGCGCTCACAGACTCAGAACAGGAAGC 2116
QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnPro 500
Db 2117 GAATGAAGTGGCTAGACGGGTCACTGTGATTACAAAACTGGAAAGCTGACAAACA 2176
QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTrpAlaGly 520
Db 2177 GATACTGGGCGAGTGGCCATGGCCAGGAGAAAGACTGTGCTGGCTTGATTTACGCAGGA 2236
QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
Db 2237 CAGTGAATGACTTCCAGTGTGATGAATCAATACTTCTTGTGAGAAGAAAGGAG 2296
QY 541 ThrValLeuSerSerAlaLeu 547
Db 2297 GCAGTACCATCATCCATATTA 2317

RESULT 5
LOCUS AB038519 3291 bp mRNA linear ROD 18-JAN-2002
DEFINITION Mus musculus src1 mRNA for scavenger receptor with C-type lectin,
complete cds.
ACCESSION AB038519
VERSION AB038519.1 GI:18146951
KEYWORDS Mus musculus cDNA to mRNA.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Nakamura, K., Funakoshi, H., Tokunaga, F. and Nakamura, T.
TITLE Molecular cloning of a mouse scavenger receptor with C-type lectin
(SRCL)(1), a novel member of the scavenger receptor family
JOURNAL Biochim. Biophys. Acta 1522 (1), 53-58 (2001)
MEDLINE 21575692
REFERENCE 2 (bases 1 to 3291)
AUTHORS Nakamura, K. and Nakamura, T.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical
School, Division of Biochemistry, Biomedical Research Center; 2-2
Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail: knakamura@onbich.med.osaka-u.ac.jp,
Tel:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)
FEATURES
source 1..3291
/organism="Mus musculus"
/db_xref="taxon:10090"
gene 1..3291
/feature="src1"
CDS 77..2305
/gene="src1"
/codon_start=1
/product="scavenger receptor with C-type lectin"
/protein_id="BAB82497.1"
/db_xref="GI:18146952"
/translation="MKDDFAEEVEQSFYKRFGEHGTCTKCIINNALKFISVLLY
ILCALLTITVAILGKYVVEKMDNVDGNETSHQYDNKLTAVESDLKLGDOAGKAL
STNSELSTFRSDILRLQLOTEIKTSKNDLLEKQAGDSLVDROSQDKETLKN
SFLTITVANKTLQAYNYTNLQDDTNVLOGNLQSQVSVIMNLNLTVQOORN
LISNLQSQVDDTSLAIQRKNDQFLQAKKDTDLKQVSLQTLAANSALA
KANNDTLEDMSOLFSSFTGMDNTITISQANEQSLKDLQDKHDTENRTAVKFSLEE
RFOVFETDIVNIISNIVTAHLRLTSLNLDVMTCTDTLFRHTDILTSLNLTVNI
RLDSLSLRMOODMRSKLDTEVANLSVYVEEMKLVDSKHGLIKNFTILQPPGPRGP
KDRGSQGPPTGNKGKGEKGPPEPPGPPAGEKRTIGTPVGPGERGSKGSGQPK
```

01234567891011121314151617181920212223242526272829303132333435363738394041424344454647484950515253545556575859606162636465666768697071727374757677787980818283848586878889909192939495969798991001011021031041051061071081091101111121131141151161171181191201211221231241251261271281291301311321331341351361371381391401411421431441451461471481491501511521531541551561571581591601611621631641651661671681691701711721731741751761771781791801811821831841851861871881891901911921931941951961971981992002012022032042052062072082092102112122132142152162172182192202212222232242252262272282292302312322332342352362372382392402412422432442452462472482492502512522532542552562572582592602612622632642652662672682692702712722732742752762772782792802812822832842852862872882892902912922932942952962972982993003013023033043053063073083093103113123133143153163173183193203213223233243253263273283293303313323333343353363373383393403413423433443453463473483493503513523533543553563573583593603613623633643653663673683693703713723733743753763773783793803813823833843853863873883893903913923933943953963973983994004014024034044054064074084094104114124134144154164174184194204214224234244254264274284294304314324334344354364374384394404414424434444454464474484494504514524534544554564574584594604614624634644654664674684694704714724734744754764774784794804814824834844854864874884894904914924934944954964974984995005015025035045055065075085095105115125135145155165175185195205215225235245255265275285295305315325335345355365375385395405415425435445455465475485495505515525535545555565575585595605615625635645655665675685695705715725735745755765775785795805815825835845855865875885895905915925935945955965975985996006016026036046056066076086096106116126136146156166176186196206216226236246256266276286296306316326336346356366376386396406416426436446456466476486496506516526536546556566576586596606616626636646656666676686696706716726736746756766776786796806816826836846856866876886896906916926936946956966976986997007017027037047057067077087097107117127137147157167177187197207217227237247257267277287297307317327337347357367377387397407417427437447457467477487497507517527537547557567577587597607617627637647657667677687697707717727737747757767777787797807817827837847857867877887897907917927937947957967977987998008018028038048058068078088098108118128138148158168178188198208218228238248258268278288298308318328338348358368378388398408418428438448458468478488498508518528538548558568578588598608618628638648658668678688698708718728738748758768778788798808818828838848858868878888898908918928938948958968978988999009019029039049059069079089099109119129139149159169179189199209219229239249259269279289299309319329339349359369379389399409419429439449459469479489499509519529539549559569579589599609619629639649659669679689699709719729739749759769779789799809819829839849859869879889899909919929939949959969979989991000100110021003100410051006100710081009101010111012101310141015101610171018101910201021102210231024102510261027102810291030103110321033103410351036103710381039104010411042104310441045104610471048104910501051105210531054105510561057105810591060106110621063106410651066106710681069107010711072107310741075107610771078107910801081108210831084108510861087108810891090109110921093109410951096109710981099110011011102110311041105110611071108110911101111111211131114111511161117111811191120112111221123112411251126112711281129113011311132113311341135113611371138113911401141114211431144114511461147114811491150115111521153115411551156115711581159116011611162116311641165116611671168116911701171117211731174117511761177117811791180118111821183118411851186118711881189119011911192119311941195119611971198119912001201120212031204120512061207120812091210121112121213121412151216121712181219122012211222122312241225122612271228122912301231123212331234123512361237123812391240124112421243124412451246124712481249125012511252125312541255125612571258125912601261126212631264126512661267126812691270127112721273127412751276127712781279128012811282128312841285128612871288128912901291129212931294129512961297129812991300

TITLE Molecular cloning and functional characterization of a human scavenger receptor with C-type lectin (SRCL), a novel member of a Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)

JOURNAL 21092718

MEDLINE 2 (bases 1 to 4330)

REFERENCE Nakamura,K. and Nakamura,T.

AUTHORS Direct Submission

JOURNAL Submitted (04-DEC-2000) Kenji Nakamura, Osaka University Graduate School of Medicine, Division of Biochemistry, Biomedical Research Center; 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

(E-mail:knakamura@bich.med.osaka-u.ac.jp, Tel:81-6-6879-3783, Fax:81-6-6879-3789)

FEATURES Location/Qualifiers

source 1..4330

gene /organism="Homo sapiens"

CDS /db_xref="taxon:9606"

1..4330

3..1871

/gene="SRCL"

/codon_start=1

/product="scavenger receptor with C-type lectin type II"

/protein_id="BAB39148.1"

/db_xref="GI:13365553"

/translation="MKDDFAEEEEVPSFGRFRIHGTQCPKHNWALKFSIILY ILCALLTIVAILGYKVVYKMDNVGTGMSRTQYDDKLTAVALDSKLKLGDTGKKAI STNSLSLFRSDILDLROQLREITKTSKNDLTLEKLOAGDALVDLRQSLKLENN SFLITVNTKTLQAYNYYVTLNQDTSVLQGNLONOMYSHNVVIMNLNLTVOQRN LITNQRSVDTTSQAIQRKNDQFNLOQVFLQAKDQDLKVKQVSLQTLAANSALA KANNQTLSDMSLSNFTGQMNITTSIOANEQNLKDLQDLKDNRTAIFKNOLEE RFLPFTDIINISYTAHMLRTLTLNLENEVRTDTTLTKHDDTLNLNLTIANI RLDSVLRMQDLMRSRLDTEVANLSVIMEEMKLVDSRHGLIKNFTILQPPGPRGP REDSGSQPPQDTGKNGQKEGEPGPPGAGERPPIGAPGPPGGRGSKSGSGPK GSRGSPGPGPGSGDPGPPGPKGELPGQPGPGFQGLQGTGVEPGVPGPLPG LPGAEGPGPGPGPPGPPGSAVPLAQLNEPTPAPEDNSKSKFSLQPGGGSACA"

BASE COUNT 1186 a 1065 c 1035 g 1044 t

ORIGIN

Alignment Scores:

Pred. No.: 9,25e-86 Length: 4330

Score: 2693.00 Matches: 547

Percent Similarity: 50.70% Conservative: 0

Best Local Similarity: 50.70% Mismatches: 0

Query Match: 90.70% Indels: 532

DB: 9 Gaps: 1

US-09-763-712A-2 (1-547) x AB052103 (1-4330)

QY 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20

DB 588 ATGTATCTCATAAATGTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAG 647

QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40

DB 648 CAGAGGAACCTCATCAGAAATCTGACCGGTCTGTGGATGACACAGCCAGGCTATCCAG 707

QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60

DB 708 CGAATCAAGAACGACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGAGGACAG 767

QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80

DB 768 GATTGGCTGAAGGAGAAAGTGAAGCTTGCAGACACTGGCTGCCAACAACTCTCGGTG 827

QY 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100

DB 828 GCCAAGCCCAACACACACACCTGGAGGATATGACAGCCAGCTCAACTCATTCACAGT 887

QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120

DB 888 CAGATGGAGAACATCACCACCTATCTCAAGCCACAGCAGCAGACCTGAAAGACCTGCAG 947

QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluGluArg 140

DB 948 GACTTACACAAGAGATGACAGAGATAGAACAGGCATCAAGTTCAACCACTGGGAGAACGC 1007

QY 141 PheGlnLeuPheGluThrAspIleValAsnIleLysSerAsnIleSerThrAlaHis 160

DB 1008 TTCCAGCTCTTTGAGACGGATATTGTGAACATCATTAGCAATATAGTTACACAGCCAC 1067

QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180

DB 1068 CACCTGCGGACGCTGACACCAATCTAATAGTACAGACCACTTGCACAGATACCTT 1127

QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200

DB 1128 ACCAAACACACAGATGATCTGACCTCTTGAATAATACCTGGCCAACATCCGTTTGGAT 1187

QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220

DB 1188 TCTGTTTCTCTCAGGATGCAACAAGATTGTGAGGTCCAGGTTAGACACTGAAGTAGCC 1247

QY 221 AsnLeuSerValIleMetGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240

DB 1248 AACTTATCAGTGTATGGAAGAAATGAGCTAGTAGACTCCAAGCATGGTCAGCTCATC 1307

QY 241 LysAsnPheThrIleLeuGlnGlyProGlyProGlyProArgGlyProArgGlyAspArgGly 260

DB 1308 AAGAATTTTACAATACTACAGGTCCACCGGGCCCCAGGGTCCAAAGAGGTGCACAGAGA 1367

QY 261 SerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGlyGlyGlyGluPro 280

DB 1368 TCCAGGGACCCCTGGCCCACTGGCAACAAGGGACAGAAAGGAGAGAGGGAGGAGCCT 1427

QY 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGlyGlu 300

DB 1428 GGACCACTTGGCCCTCGGGTGAGAGAGGCCCAATTTGGACAGCTGTCCTCCCGGAGAG 1487

QY 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320

DB 1488 CGTGGCGCAAAAGGATCTAAAGGCTCCACAGGGCCCCAAAGGCTCCCGTTCCTCCCTGGG 1547

QY 321 LysProGlyGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 340

DB 1548 AAGCCCGCCCTCAGGGCCCCAGTGGGACCCAGGGCCCCCGGGCCCCCAGGCAAGAG 1607

QY 341 GlyLeuProGlyProGlnGlyProGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360

DB 1608 GGACTCCCGCGCCCTCAGGGCCCTCCTGCTTCCAGGGACTTCAGGGCACCGCTGGGGAG 1667

QY 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380

DB 1668 CCTGGGTGCTTGGACCTCGGGGACTGCCAGGCTTGCCTGGGTACACAGGCATGCCAGGC 1727

QY 381 ProLysGlyProGlyProGlyProGlyProSerGlyValaValaProLeuAlaLeuGln 400

DB 1728 CCCAAGGGCCCCCGGCCCTCTTGCCCATCAGGAGCGGTGTGCCCTTGCCTTGCCTGCGAG 1787

QY 401 AsnGluProThrProAlaProGluAspAsn----- 410

DB 1788 AATGAGCCAAACCCGCGCAGGAGGACAAATAGTAGTCCAAAGCCCTCCCTCCAGCCAGG 1847

QY 410 ----- 410

DB 1848 GGACAGGGCAGTGCATGTGCTTTGAGACTGAGCAACCCCACTGTGTGCCCCCAACCCATG 1907

QY 410 ----- 410

DB 1908 TTATTTTGTGTTTGTGTTTGTGTTTGTGAGACAGGCTCTACTCTGTGTGCCAGGCTGGAGT 1967

QY 410 ----- 410

DB 1968 GCAGTGGCGTGATCTCGGCTGCTGCAACCCCAACATCCCGGGTTCAAGTGATTCCTCTG 2027

QY 410 ----- 410

Db 2028 COTTAGCCTCCAGTAGCTGGGAATACAAGCGTGTGCCCCACAGCCAGCTAATTTTG 2087
Qy 410 ----- 410
Db 2088 TATTTTTAGTAGAGGGTTTACAGTGTGGGAGGGTGGTTTCAAACTTCTGGCCTC 2147
Qy 410 ----- 410
Db 2148 AAGTGATCCACCTACCTAGTTTCCCAAAGTGTGGGAATACAGGTGTGAACCCACCGTT 2207
Qy 410 ----- 410
Db 2208 TTGTCCTCCCGGTCATATTATTTAAACAGCAAGGCATATAGGGGTTGGACAGGACC 2267
Qy 410 ----- 410
Db 2268 CTGCTGCCAGAACAGGCTCCATACCAGGCATGTCTCTCTTAACGCCATTTGGGTTT 2327
Qy 410 ----- 410
Db 2328 TTTAGTGTCTATTTGTTTTATGTATGTGTATGTATGGACAGGGTCTCACTCTGT 2387
Qy 410 ----- 410
Db 2388 CACCCAGGCTGGAATGACGAAGTGAAGTAGTGCAAAACATAGCTCATTTGACCT 2447
Qy 410 ----- 410
Db 2448 CCTGGCTTAGGTATCTCCCACTCAGACTCCCTCGTAGCTGGGACACAGGGGTATGC 2507
Qy 410 ----- 410
Db 2508 CACCACACCTGGCTAATTTTATTTTCTTTTGTAGAACAGGGTCTTGCTATGTTGC 2567
Qy 410 ----- 410
Db 2568 CAGGCTGGTCTGAACCTCTGGCCTCAAGCAACCCCTCCTGCCCTAGCTTTGCAAGTGC 2627
Qy 410 ----- 410
Db 2628 TGGGATCCAGGTGGAGCCACTGCACCTGGCTCTTGACACTGTTTATTTATAGAGCCA 2687
Qy 410 ----- 410
Db 2688 GGCCTCAGAAGACAAACAAATGCAAAACATTTCTTCAGGGAAGCCTGCTGGCGCCT 2747
Qy 410 ----- 410
Db 2748 GAGCGCTCTCCCTCGTTTTTACATCCTCACTCTGGAGATAGATGGGCTCCCTCGGTTAGC 2807
Qy 410 ----- 410
Db 2808 TCCATCAACATGTGGATGGAACCCCTCTGCTATGTATGACTACATTTAATTTCTTTTGAG 2867
Qy 410 ----- 410
Db 2868 AAAGGTGTAGATGTTACGGTGCCTTCTAGTCTTTTGATAATTAACCTTTGGCAAACT 2927
Qy 410 ----- 410
Db 2928 CGAATGTTTGTGTAGCAACCAAGCCTATTTTCTCTCCTTTGACCAACCGCTATTTCTAC 2987
Qy 410 ----- 410
Db 2988 TTGATTTTGGGTGAGGCCCGGGTTATTTTGTGTATCCGTGAGCCATTTTGGTGC 3047
Qy 410 ----- 410
Db 3048 TGGGTCTCACATCCAAGTGTCTGGGTTCCGTATGATTCAGACTGACGCTGAACATC 3107
Qy 410 ----- 410

Db 3108 TGTGCTGCCCTGCTTCTTGTGTTGCTTCTGTCATTTCTGTCATTTGTGAAGACGGATCC 3167
Qy 410 ----- 410
Db 3168 AAACACAGAAGNACAAAGACCCAGTCCAAGATCCATCCAGATTATGTGGTTTCGGCTTGT 3227
Qy 410 ----- 410
Db 3228 GCTCTGCAGTCCACTTGGGTGTCATGCTTGTGCGAGCACGTCGTTGAACGTGACGTGTGG 3287
Qy 410 ----- 410
Db 3288 GCCTCGTCAGGAGCGCTCAGGCTGCTCCAGCGGGACACAGTTTGGCCTCTGAAACA 3347
Qy 410 ----- 410
Db 3348 CAGCTGAAGAAAGGACAGAGTCTTCTGTGATCTTTTAACTCAATGTTTTCACATTTTGA 3407
Qy 411 -----GlyCysProHisTrpLysAsnPherThrAspLysCysTyrTyrPheSerVal 428
Db 3408 TTCTTAGGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTT 3467
Qy 429 GtLysGluLlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuVal 448
Db 3468 GAGAAAGAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAAGCTTTCACATCTTGT 3527
Qy 449 PheLleAsnThrArgGluGluGlnGlnTrpLleLysLysGlnMetValGlyArgGluSer 468
Db 3528 TTCATAACACTAGAGAGAGACAGCANTGGATATAAAACAGATGGTAGGGAGAGAGACC 3587
Qy 469 HisTrpLleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThr 488
Db 3588 CACTGGATCGGCTCACAGACTCAGAGCTGAAATGAAATGAATGGAAGTGGCTGGTGGACA 3647
Qy 489 SerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGly 508
Db 3648 TCTCCAGACTACAAAATTTGGAAGCTGGACAGCGGATAACTGGGGTCATGGCCATGGG 3707
Qy 509 ProGlyGluAspCysAlaGlyLeuLleTyrAlaGlyGlnTrpAsnAspPheGlnCysGlu 528
Db 3708 CCAGGAGAAGACTGTGCTGGGTTGATTATGCTGGCAGTGGAGCAGATTTCCAATGTGAA 3767
Qy 529 AspValAsnAsnPhleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 547
Db 3768 GACGTCAATACTTCATTTGGCAAAAGACAGGAGACAGTACTGTCATCTGCATTA 3824

RESULT 7

AX454442

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Location/Qualifiers

PAT 06-JUL-2002

Sequence 27 from Patent WO0208284.

AX454442

AX454442.1

GI:21713845

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1

Baker, K.P., Ferrara, N., Gerber, H., Gerritsen, M.E., Goddard, A.,

Goddard, A.,

Godowski, P.J., Gurney, A.L., Hillan, K.J., Marsters, S.A., Pan, J.,

Paoni, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I.

and Ye, W.

Compositions and methods for the diagnosis and treatment of

disorders involving angiogenesis

Patent: WO 0208284-A 27 31-JAN-2002;

Genentech, Inc. (US) ; Baker, Kevin P. (US) ; Ferrara, Napoleone

(US) ; Gerber, Hanspeter (US) ; Gerritsen, Mary E. (US) ; Goddard,

Audrey (US) ; Godowski, Paul J. (US) ; Gurney, Austin L. (US) ;

Hillan, Kenneth J. (US) ; Marsters, Scot A. (US) ; Pan, James (US)

; Paoni, Nicholas F. (US) ; Stephan, Jean-Philippe F. (US) ;

Watanabe, Colin K. (US) ; Williams, P. Mickey (US) ; Wood, William

I. (US)


```
source 1..2005
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 606 a 493 c 491 g 415 t
ORIGIN
Alignment Scores:
Pred. No.: 7.47e-86 Length: 2005
Score: 2688.00 Matches: 495
Percent Similarity: 99.60% Conservative: 2
Best Local Similarity: 99.20% Mismatches: 2
Query Match: 90.54% Indels: 1
DB: 6 Gaps: 0
US-09-763-712a-2 (1-547) x AX454442 (1-2005)
QY 49 LeuGlnGlnValPheLeuGlnAlaLysLysAspThrAspTrpLeuLysGluLysValGln 68
Db 8 GTCCGGCAGGTTTCTTCAAGCCCAAGAGGACAGGATGGCTGAAGGAGAAAGTGCAG 67
QY 69 SerLeuGlnThrLeuAlaAlaAsnSerAlaLeuAlaLysAlaAsnAspThrLeu 88
Db 68 AGCTGCACACACTGGCTCCCAACACTCTGGTGGCCAAAGCCCAACACGACCCCTG 127
QY 89 GluAspMetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrIle 108
Db 128 GAGGATATCAACAGCCAGCTCAACTCATTACAGAGTTCAGATGGAGAACATCACCCTATC 187
QY 109 SerGlnAlaAsnGluAlaAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsn 128
Db 188 TCTCAAGCCCAAGCAGAGAACCTGAAAGACCTGCAGGACTTACACAAAGATCAGAGAAAT 247
QY 129 ArgThrAlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIle 148
Db 248 AGACAGCCATCAAGTTCAACCACTGGAGGAACGCTTCCACTCTTTGAGAGCGATATT 307
QY 149 ValAsnIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsn 168
Db 308 GTGAACATCATAGCAATATCATAGTTACACAGCCCAACCTCGGACGCTGACCAACAT 367
QY 169 LeuAsnGluValArgThrCysThrAspThrLeuThrLysHisThrAspAspLeuThr 188
Db 368 CTAAATGAAGTCAGGACCACTTGCACAGATACCTT-ACCAACACACAGATATCTGACC 426
QY 189 SerLeuAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGln 208
Db 427 TCTTGAATAATACCTGCCCAACATCCGTTGGATTCTGTTCTCTCAGGATGCAACAA 486
QY 209 AspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGlu 228
Db 487 GATTTGATGAGTTCGAGGTGACACACTGAAGTAGCCAACTTATCATGATTATGGAAGAA 546
QY 229 MetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGly 248
Db 547 ATGAGCTAGTAGATCCCAAGCATGCTAGCTCATCAAGAATTTTACAAATACATAAGT 606
QY 249 ProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThr 268
Db 607 CCACCGGGCCCCAGGGGTCCAAGGTGACAGAGGATCCAGGGACCCCTCGGCCCAACT 666
QY 269 GlyAsnLysGlyGlnLysGlyGluLysGluProGlyProGlyProGlyProAlaGlyGlu 288
Db 667 GGCAACAAGGGACAGAAAGAGAGAGGGGGAGCCCTGGACCACTGGCCCTCGGGGTGAG 726
QY 289 ArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysLysGly 308
Db 727 AGAGGCCCAATTGGACCACTGGTCCCCCGAGAGCGTGGCGGCAAAAGGATCTAAAGGC 786
QY 309 SerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSer 328
Db 787 TCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCCCTGGAAAGCCCGGCCCTCAGGGCCCCAGT 846
QY 329 GlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyPro 348
```

```
Db 847 GGGGACCCAGGCCCCCGGGGCCACAGCAAAAGAGGACTCCCGGGCCCTCAGGGCCCT 906
QY 349 ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGly 368
Db 907 CCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCCTGGGGTGCCTGGACCTCGGGA 966
QY 369 LeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProPro 388
Db 967 CTGCCAGGCTTGCCCTGGGGTACCGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCCT 1026
QY 389 GlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGlu 408
Db 1027 GGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCAACCCCGCCAGCGAG 1086
QY 409 AspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerVal 428
Db 1087 GACAATAGCTGCCCGCCCTCACTGGGAAGAACTTTCACAGACAAATGCTACTATTTTTCAGTT 1146
QY 429 GluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuVal 448
Db 1147 GAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGGAAGCAACGCTTTCACATCTGTT 1206
QY 449 PheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValClyArgGluSer 468
Db 1207 TTCTATAACACTAGAGAGAACAGCAATGGATAAAAAAACAGATGTTAGGGAGAGAGC 1266
QY 469 HisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThr 488
Db 1267 CACTGGATCGGCCTCACAGACTCAGAGCGTGAATGAAATGAATGAAGTGGCTGGATGGACA 1326
QY 489 SerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGly 508
Db 1327 TCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGTGCATGCCATGG 1386
QY 509 ProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGlu 528
Db 1387 CCAGGAGAGAGACTGTGCTGGTGTGATTATGCTGGCGACGTGGAGCAATTTCCAATGTGAA 1446
QY 529 AspValAsnAsnPheIleCysGlyLysAspArgGluThrValLeuSerSerAlaLeu 547
Db 1447 GACGTCATTAACCTTCAATTTGCCAAAAAGACAGGAGACAGTACTGTCTCATCTGCATTA 1503
RESULT 8
AX490920 2005 bp DNA linear PAT 16-AUG-2002
LOCUS AX490920
DEFINITION Sequence 27 from Patent WO0200690.
ACCESSION AX490920
VERSION AX490920.1 GI:22323797
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Baker K.P., Ferrara N., Gerber H., Gerritsen M.E., Goddard A.,
Godowski P.J., Gurney A.L., Hillan K.J., Marsters S.A., Pan J.,
Paoni N.F., Stephan J.P., Watanabe C.K., Williams P.M., Wood W.I.
and Ye W.
TITLE Compositions and methods for the diagnosis and treatment of
disorders involving angiogenesis
JOURNAL Patent: WO 0200690-A 27 03-JAN-2002;
Genentech, Inc. (US)
FEATURES
source 1..2005
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 606 a 493 c 491 g 415 t
ORIGIN
Alignment Scores:
Pred. No.: 7.47e-86 Length: 2005
Score: 2688.00 Matches: 495
```


| | | | |
|------------|--|--|--------|
| Db | 967 | CTGCCAGGCTTGCTGGGGTACCAGGCATGCCAGCGCCCAAGAGGCCCGCCCTCT | 1026 |
| QY | 389 | GlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGlu | 408 |
| Db | 1027 | GGCCCATCAGAGGGTGGTGGCCCTGCAGAAATGAGCCACCCCGCAGCGAG | 1086 |
| QY | 409 | AspAsnGlyCysProProHisTrpLysAsnPhetrAspLysCysTyTrPheSerVal | 428 |
| Db | 1087 | GACAATAGTCGCCCGCTCACTGCAAGAACTTCACAGACAAATGCTACTATTTTCAGTT | 1146 |
| QY | 429 | GluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuVal | 448 |
| Db | 1147 | GAGAAAGAAATTTTGGAGGATGCAAGCTTTCTGTGAAGACAAAGTCTTCACATCTTGT | 1206 |
| QY | 449 | PheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSer | 468 |
| Db | 1207 | TTCTATAACACTAGAGAGAACACACATGGATAAAAAACAGATGGTAGGAGAGAGAGC | 1266 |
| QY | 469 | HisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThr | 488 |
| Db | 1267 | CACGTGGATCGCCCTCACAGACTCAGAGCGTGAAATGAATGAAAGTGGCTGGATGGGACA | 1326 |
| QY | 489 | SerProAspTyrLysAsnTrpLysAlaGlyClnProAspAsnTrpGlyHisGlyHisGly | 508 |
| Db | 1327 | TCTCCAGACTACAAAAATTTGAAAGCTGGACAGCGGATAACTGGGGTCATGGCGCATGGG | 1386 |
| QY | 509 | ProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGlu | 528 |
| Db | 1387 | CCAGGAGAAGACTGTGCTGGGTGATTATGCTGGGAGTGGACGATTTCCCATGTGAA | 1446 |
| QY | 529 | AspValAsnAsnPhelIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu | 547 |
| Db | 1447 | GACGTCAATACTTCATTTGCGAAAGACAGAGGAGACAGTACTGTCACTGTCATTA | 1503 |
| RESULT | 9 | | |
| LOCUS | HS0802985 | | |
| DEFINITION | Homo sapiens mRNA; cDNA DKFP547G1215 (from clone DKFP547G1215). | 1886 bp | linear |
| ACCESSION | AL713657 | | |
| VERSION | AL713657.1 | GI:19584339 | |
| KEYWORDS | human. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| AUTHORS | 1 (bases 1 to 1886) | | |
| TITLE | Koehler K., Beyer A., Mewes H.W., Weil B. and Wiemann S. | | |
| JOURNAL | Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY | | |
| COMMENT | Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical Research Center at the Charite, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFP547G1215) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/ Location/Qualifiers | | |
| FEATURES | | | |
| source | 1..1886 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /map="18p11.3" | | |
| | /clone="DKFP547G1215" | | |
| | /tissue_type="brain" | | |
| | /clone_lib="547 (synonym: hfbr1). Vector pSport1; host DH10B; sites NotI + SalI" | | |
| | /dev_stage="fetal" | | |
| gene | 1..1886 | | |
| | /gene="DKFP547G1215" | | |
| CDS | <1..1256 | | |

```

/gene="DKF2p547G1215"
/Note="similarity to asialoglycoprotein receptor (Mus musculus)"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAD28466.1"
/db_xref="GI:19584340"
/translation="AIKENQLEERFQFETDVIINISISYTAHHLRLTLTSLNLEVRT
TCTDLTKHTDLSLNNLTANIRLDSVSRMQDLMRSRLDTEVANLSVIMEDMKLV
DSKHGLIKNTIILQPPGPRGDRGSGQPPGTGNKGOKGKGPBPAGPAGERG
PIGPAGPGEGRGSGKSGSPGKSGSPGPGQSSGDPGPGPKGKGLPGQGP
PGFQGLQGTGEPGPGRGLPGLPGVPGMCPGPGPGPGPSGAVPLAQLNBPPTA
PEDGCPPHWKNFTDKYIFSEVEKEIFEDAKLFCEDKSSHLVFINTREQQWIKRQW
GRESHWIGLTDSERENKWLDTGSPDYKNKAGQPDNMGHGPEDCAGLIYAGOW
NDFQCEDYNNFTICEKRETVLSSAL"
polyA_signal
1838..1843
/gene="DKF2p547G1215"
1869
polyA_site
/gene="DKF2p547G1215"
569 a 444 c 445 g 428 t
BASE COUNT
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1,02e-72 Length: 1886
Score: 2309.00 Matches: 416
Percent Similarity: 99.76% Conservative: 0
Best Local Similarity: 99.76% Mismatches: 1
Query Match: 77.77% Indels: 0
DB: 9 Gaps: 0

```

US-09-763-712a-2 (1-547) x HSM802985 (1-1886)

```

QY 131 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 150
Db 3 GCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAGCTCTTTGAGAGGGATATTGTGAAC 62
QY 151 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 170
Db 63 ATCAATAGCAATATCATGTTTACAGACCCACCCAGCTGCGGAGCGCTGACCAATCAAT 122
QY 171 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 190
Db 123 GAAGTCAGGACCACTTGCACATACCTTACCAACACACACAGATGATCTGACCTCTTG 182
QY 191 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnAspLeu 210
Db 183 AATAATACCTCGCCAAACATCCGTTGGATTCTGTTCTCTCAGGATGCAACAAGATTG 242
QY 211 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 230
Db 243 ATGAGGTCAGGTTAGACACTGAAGTAGCAACTATCATGATTATGGAAGAAATGAAG 302
QY 231 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 250
Db 303 CTAGTAGACTCCAGCATGGTCAGCTCATCAAGAAATTTTACAAATACTACAAGTCCACCG 362
QY 251 GlyProArgGlyProArgGlyAspArgGlySerGlnGlnGlyProGlyProThrGlyAsn 270
Db 363 GGCCCCAGGGTCCAGAGAGTGCACAGAGGATCCAGGGACCCCTGGCCCACTGGCAAC 422
QY 271 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGly 290
Db 423 AAGGGACAGAAAGGAGAGAGGGAGCGCTGGACCACTGGCCCTCGGGGTGAGAGAGGC 482
QY 291 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGln 310
Db 483 CCAATTGGACCACTGGTCCCCCGGAGAGCGTGGCGGCAAGGATCTAAAGGCTCCACAG 542
QY 311 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlnGlyProSerGlyAsp 330
Db 543 GGCCCCAAGAGCTCCCGTGGTTCCCTCGGAAGCCCGGCCCTCAGGCTCCAGTGGGAC 602
QY 331 ProGlyProProGlyProProGlyLysGlyGluLeuProGlyProGlnGlnGlyProGly 350

```

```

Db 603 CCAGCCCCCGGGGCCACCAGGCAAGAGGAGACTCCCGGGCCCTCAGGGCCCTCTGTC 662
QY 351 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 370
Db 663 TTCCAGGAGACTTCAGGGACCCCTGGGAGCGCTGGGGTGCCTGGACCTCGGGGACTGCCA 722
QY 371 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 390
Db 723 GGCTTGCCTGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCTGGCCCA 782
QY 391 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 410
Db 783 TCAGAGCGGTGGTCCCTGGCCCTGCAGAAATGAGCCCAACCCCGCACCGGAGGACAA 842
QY 411 GlyCysProProHisTriPheAsnPheThrAspLysCysTyrTyrPheSerValGluLys 430
Db 843 GGCTGCCCGCTCACTGGAGAACTTCACAGACAAATGCTACTATTTTCAGTTGAGAAA 902
QY 431 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIle 450
Db 903 GAAATTTTGGAGGATGCAAGCTTTTCTGTGAAGACAAGCTTTCACATCTCTGTTTCATA 962
QY 451 AsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 470
Db 963 AACACTAGAGAGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGG 1022
QY 471 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 490
Db 1023 ATCGGCTCACAGACTCAGAGCGTGAAATGAATGGAAGTGGCTGGATGGGACATCTCCA 1082
QY 491 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 510
Db 1083 GACTACAAAAATTTGAAAAGCTGGACAGCGGATAAATGGGGTCATGGCCATGGCCAGGA 1142
QY 511 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 530
Db 1143 GAAGACTGTGCTGGGTGATTATCTGGGCAGTGGAAAGATTTCCAATGTGAAGACGTC 1202
QY 531 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 547
Db 1203 AATACTTCATTTGGCAAAAAGACAGGAGGACAGTACTGTCTATCTGCATTA 1253
RESULT 10
AC016128/c
LOCUS
DEFINITION Homo sapiens chromosome 18 clone RP11-324G2 map 18, WORKING DRAFT
SEQUENCE, 8 unpublished pieces.
AC016128
VERSION AC016128.4 GI:10046526
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE 1 (bases 1 to 169088)
JOURNAL Homo sapiens chromosome 18, clone RP11-324G2
REFERENCE
AUTHORS 2 (bases 1 to 169088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Collins,M., Collins,S., Collamore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karacas,A., Klein,J.,
Lehocky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,

```

TITLE
 JOURNAL
 COMMENT
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 9, 2000 this sequence version replaced gi:6649269.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

 Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

 Project Information
 Center project name: L999
 Center clone name: 324_G-2

 Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 16522 bases at least Q40
 Consensus quality: 167248 bases at least Q30
 Consensus quality: 167996 bases at least Q20
 Insert size: 177000; agarose-fp
 Insert size: 168388; sum-of-ctnigs
 Quality coverage: 6.2 in Q20 bases; agarose-fp
 Quality coverage: 6.5 in Q20 bases; sum-of-ctnigs

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L999

Center clone name: 324_G-2

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 16522 bases at least Q40

Consensus quality: 167248 bases at least Q30

Consensus quality: 167996 bases at least Q20

Insert size: 177000; agarose-fp

Insert size: 168388; sum-of-ctnigs

Quality coverage: 6.2 in Q20 bases; agarose-fp

Quality coverage: 6.5 in Q20 bases; sum-of-ctnigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

* 1 17943: contig of 17943 bp in length
 * 17944 18043: gap of 100 bp
 * 18044 22424: contig of 4381 bp in length
 * 22425 22524: gap of 100 bp
 * 22525 38094: contig of 15570 bp in length
 * 38095 38194: gap of 100 bp
 * 38195 49220: contig of 11026 bp in length
 * 49221 49320: gap of 100 bp
 * 49321 66394: contig of 17074 bp in length
 * 66395 66494: gap of 100 bp
 * 66495 91692: contig of 25198 bp in length
 * 91693 91792: gap of 100 bp
 * 91793 120869: contig of 29077 bp in length
 * 120870 120969: gap of 100 bp
 * 120970 169088: contig of 48119 bp in length.
 Location/Qualifiers

FEATURES

source

1. 169088
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="18"
 /clone="RP11-324G2"
 /clone.lib="RPC1-11 Human Male BAC"
 1. 17943

misc_feature

/note="assembly_fragment"
 clone_end:SP6
 vector_side:left

misc_feature

18044..22424

misc_feature

22525..38094

misc_feature

38195..49220

misc_feature

49321..66394

misc_feature

66495..91692

misc_feature 91793..120869
 /note="assembly_fragment"
 misc_feature 120970..169088
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right
 BASE COUNT 48103 a 33537 c 34088 g 52660 t 700 others
 ORIGIN

Alignment Scores:
 Pred. No.: 5.17e-34 Length: 169088
 Score: 1239.00 Matches: 248
 Percent Similarity: 100.00% Conservatve: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 41.73% Indels: 0
 DB: 2 Gaps: 0

US-09-763-712A-2 (1-547) x AC016128 (1-169088)

Qy 1 MetTyrSerHisAsnValIleMetAsnLeuAsnLeuThrGlnValGln 20
 |||||
 Db 165612 ATGTATTCTCATTAATGTGTCATCATCAACCTCAACAACCTGACCGAGGTGCAG 165553
 Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
 |||||
 Db 165552 CAGAGGAACCTCATCAGCAATCTGCAGCGGTCTGTGGATGACAAAGCCAGGCTATCCAG 165493
 Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysAspThr 60
 |||||
 Db 165492 CGAATCAAGACGACTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGGACACG 165433
 Qy 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80
 |||||
 Db 165432 GATTGGCTGAAGAGAAAGTGCAGAGCTGCAGACGCTGGCTGCCAACAACCTCTCGGTG 165373
 Qy 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
 |||||
 Db 165372 GCCAAAGCCCAACAGCAGACCCCTGGAGGATATGAACAGCAGCTCAACTCATTACAGGT 165313
 Qy 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 |||||
 Db 165312 CAGATGGAGAACATCACCACATCTCTCAAGCCCAAGCAGACCACTGAAAGACCTGCAG 165253
 Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
 |||||
 Db 165252 GACTTACACAAAGATGCAGAGAATAGACAGCCATCAAGTTCAACCACTGGAGGAACG 165193
 Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
 |||||
 Db 165192 TTCCAGCTCTTTGAGACGGATATTGTGAACATCATTAGCAATATCAGTTACACAGCCAC 165133
 Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 |||||
 Db 165132 CACCTGGGACCTGACCAGCAATCTAAATGAAGTCAGGACCACTTGCACATACCCCTT 165073
 Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 |||||
 Db 165072 ACCAAACACACAGATGATCTGACCTCCTTGAATAATACCTGGCCCAACATCCGTTGGAT 165013
 Qy 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
 |||||
 Db 165012 TCTGTTCTCTCAGGATGCACAAGATTGTGAGGTCCAGGTTAGACACTGAAGTAGCC 164953
 Qy 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
 |||||
 Db 164952 AACTTATCAGTGATTATGGAAGAANAATGAAGCTAGTAGACTCCACAGCATGCTCAGCTCATC 164893
 Qy 241 LysAsnPheThrIleLeuGlnGly 248
 |||||
 Db 164892 AAGAATTTTACAATACTACAAGGT 164869
 RESULT 11
 AP000900


```

/note="assembly_fragment"
149151..156105
/note="assembly_fragment"
156206..162048
/note="assembly_fragment"
162149..165118
/note="assembly_fragment clone_end:SP6 vector_side:left"
165219..167480
/note="assembly_fragment"
167581..169986
/note="assembly_fragment"
170087..171909
/note="assembly_fragment"
172010..172317
/note="assembly_fragment clone_end:T7 vector_side:right"
172418..174663
/note="assembly_fragment"
174764..176271
/note="assembly_fragment"
176372..178022
/note="assembly_fragment"
BASE COUNT 53972 a 34911 c 36303 g 50935 t 1901 others
ORIGIN
Alignment Scores:
Pred. No.:      5.38e-34      Length:      178022
Score:          1239.00      Matches:      248
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      41.73%      Indels:      0
DB:              2          Gaps:      0

US-09-763-712a-2 (1-547) x AP000900 (1-178022)
QY 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
Db 13158 ATGTATTCATTAATGTGTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTCAG 13217
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
Db 13218 CAGAGGAACCTCATCAGGAATCTGCAGCGTCTGTGGATGACACAGCCAGGCTATCCAG 13277
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
Db 13278 CGAATCAAGAACGACTTCAAAATCTGCAGCAGGTTTCTTCAAGCCAAAGACACAGC 13337
QY 61 AspTyrLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
Db 13338 GATTGGCTGAAGAGAAAGTCAGAGCTTCAGACGCTGCGCTGCCAACCAACTCTGCGTTG 13397
QY 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
Db 13398 GCCAAAGCCAAACAGACACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGT 13457
QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluAlaAsnLeuLysAspLeuGln 120
Db 13458 CAGATGGAGAAATCACCACACTCTCTCAAGCCAAAGCAGACAGAACCTGAAGACCTGCAG 13517
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
Db 13518 GACTTACACAAAGATGCAGAGAAATAGACAGCCATCAAGTTCAACCACTGGAGGAACCC 13577
QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
Db 13578 TTCCAGCTCTTTGAGACGGATATTGGAACATCATAGCAATATCAGTATACACGCCAC 13637
QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
Db 13638 CACCTCGGAGCCTGACCAGCAATCTAAATGAATGAGTCAGGACCCACTTGCACGATACCC 13697
QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
Db 13698 ACCAAACACACAGATGATGTGACCTCTTGAATAATACCTTGCCCAACATCCGTTTGGAT 13757

```

```

QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
Db 13758 TCTGTTTCTCTCAGGATGCAACAAGATTGTGATCAGGTTCAGACTGAAGTAGCC 13817
QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLeu 240
Db 13818 AACTTATCAGTATTATGGAAGAAATGAAGCTAGTAGCTCAAGCATGCTCAGCTCATC 13877
QY 241 LysAsnPheThrIleLeuGlnGly 248
Db 13878 AAGAATTTTACAATACTACAGGT 13901

RESULT 12
AP001022
LOCUS      187635 bp      DNA      linear      HTG 30-MAY-2000
DEFINITION DRAFT SEQUENCE, 33 unordered pieces.
ACCESSION AP001022
VERSION    AP001022.2 GI:8117692
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-815L4.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 187635)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 187,635 genomic DNA of RP11.3
Published Only in Database
2 (bases 1 to 187635)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (05-JAN-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6997772.
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: Humdraft18
Center clone name: RP11-815L4
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167192 bases at least Q40
Consensus quality: 177020 bases at least Q30
Consensus quality: 182105 bases at least Q20
Insert size: 184435; sum-of-contigs
Quality coverage: 4.35x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
33 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1
21332 contig of 21332 bp in length
21433 42072 contig of 20640 bp in length
42173 54336 contig of 12164 bp in length
54437 66766 contig of 12330 bp in length
66867 77849 contig of 10983 bp in length
77950 87798 contig of 9849 bp in length
87899 99510 contig of 11612 bp in length

```

99611 107273 contig of 7663 bp in length
 107374 113189 contig of 5816 bp in length
 113290 120559 contig of 7270 bp in length
 120660 126468 contig of 5809 bp in length
 126569 131619 contig of 5051 bp in length
 131720 136810 contig of 5091 bp in length
 136911 139925 contig of 3015 bp in length
 140026 144232 contig of 4207 bp in length
 144333 148902 contig of 4570 bp in length
 149003 152918 contig of 3916 bp in length
 153019 156399 contig of 3381 bp in length
 156500 158933 contig of 2434 bp in length
 159034 161885 contig of 2851 bp in length
 161985 164301 contig of 2317 bp in length
 164402 166300 contig of 1899 bp in length
 166401 168580 contig of 2180 bp in length
 168681 171270 contig of 2590 bp in length
 171371 174092 contig of 2722 bp in length
 174193 176095 contig of 1903 bp in length
 176196 177890 contig of 1695 bp in length
 177991 179786 contig of 1796 bp in length
 179887 181478 contig of 1592 bp in length
 181579 182895 contig of 1317 bp in length
 182996 184841 contig of 100 bp in length
 184942 186345 contig of 1404 bp in length
 186446 187635 contig of 1190 bp in length

Sequence updated (26-May-2000).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 33 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 21332: contig of 21332 bp in length
 21333 21432: gap of 100 bp
 21433 42072: contig of 20640 bp in length
 42073 42172: gap of 100 bp
 42173 54336: contig of 12164 bp in length
 54337 54436: gap of 100 bp
 54437 66766: contig of 12330 bp in length
 66767 68666: gap of 100 bp
 68667 77849: contig of 10983 bp in length
 77850 77949: gap of 100 bp
 77950 87798: contig of 9849 bp in length
 87799 87898: gap of 100 bp
 87899 99510: contig of 11612 bp in length
 99511 99610: gap of 100 bp
 99611 107273: contig of 7663 bp in length
 107274 107373: gap of 100 bp
 107374 113189: contig of 5816 bp in length
 113190 113289: gap of 100 bp
 113290 120559: contig of 7270 bp in length
 120560 120659: gap of 100 bp
 120660 126468: contig of 5809 bp in length
 126469 126568: gap of 100 bp
 126569 131619: contig of 5051 bp in length
 131620 131719: gap of 100 bp
 131720 136810: contig of 5091 bp in length
 136811 136910: gap of 100 bp
 136911 139925: contig of 3015 bp in length
 139926 140025: gap of 100 bp
 140026 144232: contig of 4207 bp in length
 144233 144332: gap of 100 bp
 144333 148902: contig of 4570 bp in length
 148903 149002: gap of 100 bp
 149003 152918: contig of 3916 bp in length
 152919 153018: gap of 100 bp
 153019 156399: contig of 3381 bp in length
 156400 156499: gap of 100 bp
 156500 158933: contig of 2434 bp in length
 158934 159033: gap of 100 bp

* 159034 161884: contig of 2851 bp in length
 * 161885 161984: gap of 100 bp
 * 161985 164301: contig of 2317 bp in length
 * 164302 164401: gap of 100 bp
 * 164402 166300: contig of 1899 bp in length
 * 166301 166400: gap of 100 bp
 * 166401 168580: contig of 2180 bp in length
 * 168581 168680: gap of 100 bp
 * 168681 171270: contig of 2590 bp in length
 * 171271 171370: gap of 100 bp
 * 171371 174092: contig of 2722 bp in length
 * 174093 174192: gap of 100 bp
 * 174193 176095: contig of 1903 bp in length
 * 176096 176195: gap of 100 bp
 * 176196 177890: contig of 1695 bp in length
 * 177891 177990: gap of 100 bp
 * 177991 179786: contig of 1796 bp in length
 * 179787 179886: gap of 100 bp
 * 179887 181478: contig of 1592 bp in length
 * 181479 181578: gap of 100 bp
 * 181579 182895: contig of 1317 bp in length
 * 182896 182995: gap of 100 bp
 * 182996 184841: contig of 1846 bp in length
 * 184842 184941: gap of 100 bp
 * 184942 186345: contig of 1404 bp in length
 * 186346 186445: gap of 100 bp
 * 186446 187635: contig of 1190 bp in length.

FEATURES
 source

1..187635
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18p11.3"
 /clone="Rp11-815L4"

misc_feature
 1..21332
 /note="assembly_fragment"
 misc_feature
 21433..42072
 /note="assembly_fragment"
 misc_feature
 42173..54336
 /note="assembly_fragment"
 misc_feature
 54437..66766
 /note="assembly_fragment"
 misc_feature
 66867..77849
 /note="assembly_fragment"
 misc_feature
 77950..87798
 /note="assembly_fragment clone_end:T7 vector_side:right"
 misc_feature
 87899..99510
 /note="assembly_fragment"
 misc_feature
 99611..107273
 /note="assembly_fragment"
 misc_feature
 107374..113189
 /note="assembly_fragment clone_end:SP6 vector_side:left"
 misc_feature
 113290..120559
 /note="assembly_fragment"
 misc_feature
 120660..126468
 /note="assembly_fragment"
 misc_feature
 126569..131619
 /note="assembly_fragment"
 misc_feature
 131720..136810
 /note="assembly_fragment"
 misc_feature
 136911..139925
 /note="assembly_fragment"
 misc_feature
 140026..144232
 /note="assembly_fragment"
 misc_feature
 144333..148902
 /note="assembly_fragment"
 misc_feature
 149003..152918
 /note="assembly_fragment"
 misc_feature
 153019..156399

Alignment Scores:
 Pred. NO.:
 Score: 5.61e-34 Length: 187635
 1239.00 Matches: 248

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 41.73% Indels: 0
 DB: 2 Gaps: 0

US-09-763-712A-2 (1-547) x AP001022 (1-187635)

Qy 1 MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
 Db 58408 ATGTATTCATATATGTGTCATCATGAACCTCAACACCTGAACCTGACCCAGGTGCAG 58467
 Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
 Db 58468 CAGAGGAACCTCATACCAATCTGCAGCGGTCTGGAGTACACAAAGCAGGCTATCCAG 58527
 Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
 Db 58528 CGAATCAGACGACTTTCAAATCTGCAGCAGGTTCCTTCAGCCCAAGAGACACAG 58587
 Qy 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80
 Db 58588 GATTGGCTGAAGGAGAAAGTGCAGAGCTTGACAGCGCTGGCTGCCAACAACTCTCGCTTG 58647
 Qy 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
 Db 58648 GCCAAGGCCAACACGACACCTGGAGGATATGAACAGCCAGCTCAACTTCACAGGT 58707
 Qy 101 GlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 Db 58708 CAGATGGAGACATCACCACATCTCTCAGCCACAGCAGACCTGAAAGACCTGCAG 58767
 Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
 Db 58768 GACTTACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAAACCACTGGAGAACGC 58827
 Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
 Db 58828 TTCCAGCTCTTGAGACGGATATTGTGAACATCATATTAGCAATATCATGTTACACGCCAC 58887
 Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 Db 58888 CACCTGCGGAGCTGACCAAGCAATCTAATGAAGTCAGGACCACTTGACAGATACCTT 58947
 Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 Db 58948 ACCAAACACACAGATGATCTGACCTCTTGAATATACCTGGCCCAACATCCGTTTGGAT 59007
 Qy 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
 Db 59008 TCTGTTCTCTCAGGATGCAACAAGATTGTGATGAGTGCAGGTTAGACACTGAAGTAGCC 59067
 Qy 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
 Db 59068 AACTATCAGTGATTATGAGAGAATAGAGCTAGTAGACTCCACGATGGTCAGCTCATC 59127
 Qy 241 LysAsnPheThrIleLeuGlnGly 248
 Db 59128 AAGATTTTACAATACTACAAGGT 59151

RESULT 13

AP000915/c

LOCUS

DEFINITION Homo sapiens genomic DNA, chromosome 18p clone:RP11-720L2, complete sequences.

ACCESSION AP000915

VERSION AP000915.5 GI:20334314

KEYWORDS HTG.

SOURCE

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Homo sapiens genomic DNA
 Published Only in Database (1999)

2 (bases 1 to 188439)

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

Direct Submission

Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,

Tel:81-45-503-9111, Fax:81-45-503-9170)

On Apr 26, 2002 this sequence version replaced gi:9188470.

FEATURES

source

1. 188439

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="18"

/map="18p"

/clone="RP11-720L2"

BASE COUNT 54003 a 41134 c 40027 g 53275 t

ORIGIN

Alignment Scores:

Pred. No.: 5.63e-34 Length: 188439

Score: 1239.00 Matches: 248

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 41.73% Indels: 0

DB: 9 Gaps: 0

US-09-763-712A-2 (1-547) x AP000915 (1-188439)

Qy 1 MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20

Db 56742 ATGTATTCATATATGTGTCATCATGAACCTCAACACCTGAACCTGACCCAGGTGCAG 56683

Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40

Db 56682 CAGAGGAACCTCATCAGCAATCTGCAGCGGTCTGGATGACACAGCCAGCTATCCAG 56623

Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60

Db 56622 CGAATCAAGACGACTTTCAAAATCTGCAGCAGGTTCCTTCAAGCCCAAGAGGACAG 56563

Qy 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80

Db 56562 GATTGGCTGAAGGAGAAAGTGCAGAGCTTGACAGCGCTGGCTGCCAACAACTCTGCGTTG 56503

Qy 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100

Db 56502 GCCAAGGCCAACACGACACCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGT 56443

Qy 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120

Db 56442 CAGATGGAGACATCACCACATCTCTCAAGCCCAACAGCAGACCTGAAAGACCTGCAG 56383

Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140

Db 56382 GACTTACAAAGATGCAGAGAAATAGACAGCCATCAAGTTCACCAACTGGAGGAACGC 56323

Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160

Db 56322 TTCCAGCTCTTGAGACGGATATTGTGAACATCATTCAGCAATATCATGTACACGCCAC 56263

Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180

Db 56262 CACCTGCGGAGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGACAGATACCTT 56203

Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200

Db 56202 ACCAAACACACAGATGATCTGACCTCTTGAATAATACCTTGCCCAACATCCGTTTGGAT 56143

* 167067 167156: gap of 100 bp in length
* 167167 170120: contig of 2954 bp in length
* 170121 170220: gap of 100 bp in length
* 170221 172430: contig of 2210 bp in length
* 172431 172530: gap of 100 bp in length
* 172531 175598: contig of 3068 bp in length
* 175599 175698: gap of 100 bp in length
* 175699 177118: contig of 1420 bp in length
* 177119 177218: gap of 100 bp in length
* 177219 178531: contig of 1313 bp in length
* 178532 178631: gap of 100 bp in length
* 178632 179857: contig of 1226 bp in length
* 179858 179957: gap of 100 bp in length
* 179958 181064: contig of 1107 bp in length
* 181065 181164: gap of 100 bp in length
* 181165 182735: contig of 1571 bp in length
* 182736 182835: gap of 100 bp in length
* 182836 183923: contig of 1088 bp in length
* 183924 184023: gap of 100 bp in length
* 184024 185810: contig of 1787 bp in length
* 185811 185910: gap of 100 bp in length
* 185911 187145: contig of 1235 bp in length
* 187146 187245: gap of 100 bp in length
* 187246 188255: contig of 1010 bp in length.

FEATURES

source

1. .188255
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-839023"
1. .16096
/note="assembly_fragment"
16197. .33789
/note="assembly_fragment"
33890. .51010
/note="assembly_fragment"
51111. .63558
/note="assembly_fragment"
63659. .75056
/note="assembly_fragment"
75157. .84737
/note="assembly_fragment"
84838. .90909
/note="assembly_fragment"
91010. .98615
/note="assembly_fragment"
98716. .105978
/note="assembly_fragment"
106079. .113049
/note="assembly_fragment"
113150. .119201
/note="assembly_fragment"
119302. .126582
/note="assembly_fragment"
126683. .132070
/note="assembly_fragment"
132171. .137138
/note="assembly_fragment"
137239. .143055
/note="assembly_fragment"
143156. .148639
/note="assembly_fragment"
148740. .153086
/note="assembly_fragment clone_end:SP6 vector_side:right"
153187. .158921
/note="assembly_fragment"
159022. .163174

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

Alignment Scores:

Pred. No.: 4.87e-33 Length: 188255

Score: 1212.00 Matches: 247

Percent Similarity: 99.60% Conservative: 0

Best Local Similarity: 99.60% Mismatches: 1
Query Match: 40.82% Indels: 1
DB: 2 Gaps: 0

US-09-763-712a-2 (1-547) x AP000939 (1-188255)

Qy 1 MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20

Db 102661 ATGTATTCTCATATAATGTGTTCATCATGACCTCAACACCTGAACCTGACCGGTGACG 102720

Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40

Db 102721 CAGAGAACCTCATCACGAATCTGCAGCGGTCTGTGGATGACACAGCAGGCTATCCAG 102780

Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60

Db 102781 CGAATCAAGACGACTTTCAAAATCTGCAGCAGGTTTCTTCAAGCCAAAGACACG 102840

Qy 61 AspTyrLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80

Db 102841 GATTGGCTGAAGGAAAGTGCAGAGCTTGCAGAGCGTGGCTGCCAACAACTCTGCGTGG 102900

Qy 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100

Db 102901 GCCAAAGCCCAACAGCACCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGT 102960

Qy 101 GlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120

Db 102961 CAGATGGAGACATCACCACTATCTCTCAAGCCACAGCAGACAGCTGAAGACCTGCAG 103020

Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140

Db 103021 GACTTACAAAGATGCAGAGAATAGAACAGCCATCAAGTTCAACCACTGGAGGAACGC 103080

Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160

Db 103081 TTCCAGCTCTTTGAGACGATATTGTGAACATCATAGCAATATCAGTTACACAGCCAC 103140

Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180

Db 103141 CACCTGGGACGCTCACCAGCAATCTAATGAAGTCAGGACCCTTGCACATACCTTT 103200

Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200

Db 103201 ACCAAACACACAGATGATCTGACCTCCTTGAATAATACCTGGCCAAACATCCGTTT-GAT 103259

Qy 201 SerValSerLeuArgMetClnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220

Db 103260 TCTGTTTCTCTCAGGATGCACAAGATTGTGAGGTGAGGTTAGACACTGAAGTAGGCC 103319

Qy 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240

Db 103320 AACTTATCAGTGATATGGAAGAAATGAGCTAGTAGACTCCCAAGCATGTCAGCTCATC 103379

Qy 241 LysAsnPheThrIleLeuGlnGly 248

Db 103380 AAGAATTTTACAATACTACACAGGT 103403

RESULT 15

AC114677/c 182029 bp DNA linear HTG 10-JUN-2002

LOCUS Mus musculus clone RP24-213K19, WORKING DRAFT SEQUENCE, 10 ordered

DEFINITION pieces.

AC114677

AC114677 3 GI:21362159

HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 182029)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Mus musculus, clone RP24-213K19

Db 1877 ATGTATTCTCAGAGCGTGGTTATCATGAAACCTCAACAACCTGAACCTAACCAGGTTCCAG 1818
Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
Db 1817 CAGAGGAACCTTATCTCAAACTCGCAGCAGTCTGTGGATGACACAGGCTGGCCATCCAG 1758
Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnGlnValPheLeuGlnAlaLysLysAspThr 60
Db 1757 CCAATTAAGAATGATTCCAAAATCTGCAGCAGGTTTCTTCAAGCCAAGAAGCACACC 1698
Qy 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
Db 1697 GATTGGCTAAAGGAAAAAGTACAGAGCTTGCAGACATTGGCTGCCAACAACTCTGCCCTG 1638
Qy 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
Db 1637 GCCAAGCCCAACATGACACCTAGAGGATATGATAGCCAGCTCAGCTCATTCACAGGT 1578
Qy 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
Db 1577 CAGATGGACAACATTACCACATATCTCACAGGCCAACGAGCAGAGGCTGAAAAGACCTTCAG 1518
Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
Db 1517 GACTTACACAAGGATACAGAAATAGAACAGCTGTCAAGTTTCAGCCAACTTGAGGAACGC 1458
Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
Db 1457 TTCCAGGCTTTTGACACAGATATTGTGAACATCATTAGCAACATCAGCTACACGCCCAT 1398
Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
Db 1397 CACCTGAGGACACTCACCGCAATCTGAATGATGTAGGACCACATGCACACACCTTG 1338
Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
Db 1337 ACCAGACACCGGATGACCTGACCTCCTTGAATAACACACTAGTCAACATCCGCTTGGAT 1278
Qy 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
Db 1277 TCTATTCTCTCAGGATGCAGACACATGATGAGTCAAGTTAGACACTGAAGTGGCC 1218
Qy 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
Db 1217 AACTTATCAGTGGTTATGGAAGAGATGAACTGGTTGACTCCAGCAGCGTCAGCTCATC 1158
Qy 241 LysAsnPheThrIleLeuGlnGly 248
Db 1157 AGAACTTTACCATTTCTACAAAGGT 1134

Search completed: March 21, 2003, 11:12:28
Job time : 4023.46 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:02:25 ; Search time 4292.42 Seconds
(without alignments)
11492.169 Million cell updates/sec

Title: US-09-763-712A-l_COPY_l_1695

Perfect score: 1695

Sequence: 1 gtcacgaatctgcagcaaga.....cagtactgtcatctgcatta 1695

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pet.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pi.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|-----------|--------------------|
| 1 | 1695 | 100.0 | 2983 | 9 | AB005145 | AB005145 Homo sapi |
| 2 | 1693.4 | 99.9 | 2641 | 6 | AX047353 | AX047353 Sequence |
| 3 | 1691.8 | 99.8 | 3058 | 9 | AB038518 | AB038518 Homo sapi |
| 4 | 1477.6 | 87.2 | 2005 | 6 | AX454442 | AX454442 Sequence |
| 5 | 1477.6 | 87.2 | 2005 | 6 | AX490920 | AX490920 Sequence |
| 6 | 1359 | 80.2 | 2637 | 10 | AB078434 | AB078434 Mus muscu |
| 7 | 1354.2 | 79.9 | 3291 | 10 | AB038519 | AB038519 Mus muscu |
| 8 | 1281.2 | 75.6 | 4330 | 9 | AB052103 | AB052103 Homo sapi |
| 9 | 1251.4 | 73.8 | 1886 | 9 | HS0802985 | AL713657 Homo sapi |
| c 10 | 798 | 47.1 | 169088 | 2 | AC016128 | AL713657 Homo sapi |
| c 11 | 798 | 47.1 | 178022 | 2 | AF000900 | AP000900 Homo sapi |
| c 12 | 798 | 47.1 | 187635 | 2 | AP001022 | AP001022 Homo sapi |
| c 13 | 798 | 47.1 | 188439 | 9 | AP000915 | AP000915 Homo sapi |
| c 14 | 786 | 46.4 | 188255 | 2 | AP000939 | AP000939 Homo sapi |
| c 15 | 662 | 39.1 | 182029 | 2 | AC114677 | AC114677 Mus muscu |
| c 16 | 662 | 39.1 | 193208 | 2 | AC102618 | AC102618 Mus muscu |
| c 17 | 629.8 | 37.2 | 71044 | 2 | AC024368 | AC024368 Homo sapi |
| c 18 | 437.4 | 25.8 | 188255 | 2 | AP000939 | AP000939 Homo sapi |
| c 19 | 388 | 22.9 | 182029 | 2 | AC114677 | AC114677 Mus muscu |
| c 20 | 388 | 22.9 | 193208 | 2 | AC102618 | AC102618 Mus muscu |
| c 21 | 378.4 | 22.3 | 130763 | 2 | AC112416 | AC112416 Rattus no |
| c 22 | 248.2 | 14.6 | 3636 | 9 | AB007829 | AB007829 Homo sapi |
| c 23 | 248.2 | 14.6 | 3685 | 6 | E32511 | E32511 Scavenger r |
| c 24 | 248.2 | 14.6 | 3810 | 6 | E32509 | E32509 Scavenger r |
| c 25 | 222.8 | 13.1 | 2215 | 10 | BC026446 | BC026446 Mus muscu |
| c 26 | 146.8 | 8.7 | 178022 | 2 | AP000900 | AP000900 Homo sapi |
| c 27 | 146.8 | 8.7 | 187635 | 2 | AP001022 | AP001022 Homo sapi |
| c 28 | 135.6 | 8.0 | 1711 | 9 | AB007830 | AB007830 Homo sapi |
| c 29 | 135.6 | 8.0 | 1877 | 6 | E32510 | E32510 Scavenger r |
| c 30 | 134.6 | 7.9 | 95002 | 9 | AF233390 | AF233390 Homo sapi |
| c 31 | 134.6 | 7.9 | 134514 | 2 | AC025192 | AC025192 Homo sapi |
| c 32 | 134.6 | 7.9 | 157324 | 9 | AC013643 | AC013643 Homo sapi |
| c 33 | 134.6 | 7.9 | 175382 | 9 | AF235104 | AF235104 Homo sapi |
| c 34 | 132.8 | 7.8 | 152181 | 2 | AC126444 | AC126444 Mus muscu |
| c 35 | 132.8 | 7.8 | 219809 | 2 | AC125058 | AC125058 Mus muscu |
| c 36 | 129 | 7.6 | 6156 | 10 | AB009993 | AB009993 Mus muscu |
| c 37 | 128.2 | 7.6 | 2193 | 3 | PALCOLIA | M25282 P.lividus a |
| c 38 | 125.8 | 7.4 | 6076 | 10 | AF272661 | AF272661 Rattus no |
| c 39 | 124.2 | 7.3 | 1041 | 4 | AF138883 | AF138883 Bos tauru |
| c 40 | 124.2 | 7.3 | 2053 | 5 | AB008374 | AB008374 Oncorhync |
| c 41 | 124.2 | 7.3 | 4502 | 5 | AB052836 | AB052836 Oncorhync |
| c 42 | 122.6 | 7.2 | 4995 | 10 | MMU16789 | U16789 Mus musculu |
| c 43 | 121.6 | 7.2 | 810 | 6 | AR014116 | AR014116 Sequence |
| c 44 | 121.4 | 7.2 | 4628 | 4 | AB008683 | AB008683 Bos tauru |
| c 45 | 121.4 | 7.2 | 5676 | 6 | E07265 | E07265 cDNA encodi |

ALIGNMENTS

RESULT 1
AB005145
LOCUS
DEFINITION Homo sapiens-CL-p1 mRNA for collectin placenta 1, complete cds.
ACCESSION AB005145
VERSION AB005145.1 GI:17026100
KEYWORDS
SOURCE Homo sapiens female tissue; placenta cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Keshi, H., Sakai, Y., Fukuo, A., Sakamoto, T., Itabe, H., Suzutani, T., Ogasawara, M., Yoshida, I. and Wakamiya, N.
2983 bp mRNA linear PRI 21-NOV-2001

Pred. No. is the number of results predicted by chance to have a


```
QY 1501 AGTGCTGGATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCGCGGATAAC 1560
|||||
Db 2102 AGTGCTGGATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCGCGGATAAC 2161
|||||
QY 1561 TGGGGTCATGGCCATGGGCGAGGAGAGACTGTGCTGGGTGATTTATCTGGGCGAGTGG 1620
|||||
Db 2162 TGGGGTCATGGCCATGGGCGAGGAGAGACTGTGCTGGGTGATTTATCTGGGCGAGTGG 2221
|||||
QY 1621 AACGATTTCCAAATGTGAAGACGTCAATAACTTTCATTTGCGAAAGACAGGAGACAGTA 1680
|||||
Db 2222 AACGATTTCCAAATGTGAAGACGTCAATAACTTTCATTTGCGAAAGACAGGAGACAGTA 2281
|||||
QY 1681 CTGTCATCTGCATTA 1695
|||||
Db 2282 CTGTCATCTGCATTA 2296
|||||

RESULT 2
AX047353
LOCUS AX047353 2641 bp DNA linear PAT 15-DEC-2000
DEFINITION Sequence 39 from Patent WO0068380.
ACCESSION AX047353
VERSION AX047353.1 GI:11876599
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2641)
AUTHORS Bandman,O., Hillman,J.L., Tang,Y.T., Lal,P., Yue,H., Baughn,M.R.,
Lu,D.A. and Azimzai,Y.
TITLE Extracellular matrix and adhesion-associated proteins
JOURNAL -Patent: WO 0068380-A 39 16-NOV-2000;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
source
1..2641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 4586187CB1"
BASE COUNT 823 a 628 c 608 g 582 t
ORIGIN

Query Match 99.9%; Score 1693.4; DB 6; Length 2641;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1694; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCACGAATCTGCAGCAAGATACCAGCTGCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
|||||
Db 243 GTCAGAAATCTGCAGCAAGATACCAGCTGCTCCAGGGCAATCTGCAGAACCAAAATGTAT 302
|||||
QY 61 TCTCATAAATGTGTCTATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAGACAGAG 120
|||||
Db 303 TCTCATAAATGTGTCTATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAGACAGAG 362
|||||
QY 121 AACCTCATCAGCAATCTGCAGGGGTCTGTGGATGACACAGCCAGGCTATCCAGCGCAATC 180
|||||
Db 363 AACCTCATCAGCAATCTGCAGGGGTCTGTGGATGACACAGCCAGGCTATCCAGCGCAATC 422
|||||
QY 181 AAGACGACTTTCAAATCTGCAGCAGGTTTTCATCAAGCCCAAGAGACACGAGTGG 240
|||||
Db 423 AAGACGACTTTCAAATCTGCAGCAGGTTTTCATCAAGCCCAAGAGACACGAGTGG 482
|||||
QY 241 CTGAAGGAGAAAGTGCAGAGCTTGCAGACGCTGGCTGCCAACAACTCTCGGTGGCCAAA 300
|||||
Db 483 CTGAAGGAGAAAGTGCAGAGCTTGCAGACGCTGGCTGCCAACAACTCTCGGTGGCCAAA 542
|||||
QY 301 GCCAACACGACACCCCTGGAGGATATGAACGCCAGCTCAACTCAATTCACAGGTTCAGATG 360
|||||
Db 543 GCCAACACGACACCCCTGGAGGATATGAACGCCAGCTCAACTCAATTCACAGGTTCAGATG 602
|||||
QY 361 GAGACATCACCACATATCTCTAAGCCACAGCAGCAGAACCTGAAGACCTGCAGGACTTA 420
|||||
Db 603 GAGACATCACCACATATCTCTAAGCCACAGCAGCAGAACCTGAAGACCTGCAGGACTTA 662
|||||
```

```
QY 421 CACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTTCCAG 480
|||||
Db 663 CACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTTCCAG 722
|||||
QY 481 CTCCTTTGACAGCGATATTGTGAACATCATTAAGCAATATCAGTTACACAGCCACACCTG 540
|||||
Db 723 CTCCTTTGACAGCGATATTGTGAACATCATTAAGCAATATCAGTTACACAGCCACACCTG 782
|||||
QY 541 CGSACGCTGACACAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCTTTACCAA 600
|||||
Db 783 CGSACGCTGACACAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCTTTACCAA 842
|||||
QY 601 CACACAGATGATCTGACCTCTCTTGAATAATACCTCGCCCAACATCCGTTTGGATTTCTGTT 660
|||||
Db 843 CACACAGATGATCTGACCTCTCTTGAATAATACCTCGCCCAACATCCGTTTGGATTTCTGTT 902
|||||
QY 661 TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTAGACACTGAAGTAGGCAACTTA 720
|||||
Db 903 TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTAGACACTGAAGTAGGCAACTTA 962
|||||
QY 721 TCAGTGATTTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGCTCAGCTCATCAAGAAT 780
|||||
Db 963 TCAGTGATTTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGCTCAGCTCATCAAGAAT 1022
|||||
QY 781 TTTACAATACTACAAGGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCCCAG 840
|||||
Db 1023 TTTACAATACTACAAGGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCCCAG 1082
|||||
QY 841 GGACCCCTTGGCCCACTGGCAACAAGGACAGAAAGAGAGAGAGGGGAGCGCTTGACCA 900
|||||
Db 1083 GGACCCCTTGGCCCACTGGCAACAAGGACAGAAAGAGAGAGAGGGGAGCGCTTGACCA 1142
|||||
QY 901 CTGCGCCCTGCGGGTCAGAGAGGCCCAATTTGGACCACTGCTCCCGCCCGAGAGCGTGGC 960
|||||
Db 1143 CTGCGCCCTGCGGGTCAGAGAGGCCCAATTTGGACCACTGCTCCCGCCCGAGAGCGTGGC 1202
|||||
QY 961 GGCAAAAGGATCTAAAGGCTCCACAGGGCCCCAAAGGCTCCCGTGTTCCTCGGAAAGCCC 1020
|||||
Db 1203 GGCAAAAGGATCTAAAGGCTCCACAGGGCCCCAAAGGCTCCCGTGTTCCTCGGAAAGCCC 1262
|||||
QY 1021 GGCCCTCAGGGCCCCAGTGGGGACCCAGAGCCCCCGGGCCCCACAGGCAAAAGAGGACTC 1080
|||||
Db 1263 GGCCCTCAGGGCCCCAGTGGGGACCCAGAGCCCCCGGGCCCCACAGGCAAAAGAGGACTC 1322
|||||
QY 1081 CCGGGCCCTCAGGGCCCTCTGCTGCTCCAGGGACTTCAGGGCACCTTGGGGAGCGCTGGG 1140
|||||
Db 1323 CCGGGCCCTCAGGGCCCTCTGCTGCTTCACAGGGACTTCAGGGCACCTTGGGGAGCGCTGGG 1382
|||||
QY 1141 GTGCTCTGGACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAAGGCTGCCAGGCCCCCAAG 1200
|||||
Db 1383 GTGCTCTGGACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAAGGCTGCCAGGCCCCCAAG 1442
|||||
QY 1201 GGCCCCCGGGCCCTCTGCGCCCATCAGAGCGGTGGTGGCCCTTGGCCCTGCGAGATGAG 1260
|||||
Db 1443 GGCCCCCGGGCCCTCTGCGCCCATCAGAGCGGTGGTGGCCCTTGGCCCTGCGAGATGAG 1502
|||||
QY 1261 CCAACCCCGGACCGGAGGACAAATGCTGCCCGCTCACTTGAAGAACTTCACAGACAAA 1320
|||||
Db 1503 CCAACCCCGGACCGGAGGACAAATGCTGCCCGCTCACTTGAAGAACTTCACAGACAAA 1562
|||||
QY 1321 TGTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCTTTTCTGTGAAGAC 1380
|||||
Db 1563 TGTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCTTTTCTGTGAAGAC 1622
|||||
QY 1381 AGCTCTTCACATCTTTGTTTCAATAACACTAGAGAGAACAGCAATGGATATAAACAACAG 1440
|||||
Db 1623 AGCTCTTCACATCTTTGTTTTCATAAACAACACTAGAGAGAACAGCAATGGATATAAACAACAG 1682
|||||
QY 1441 ATGGTAGGAGAGAGAGCGCACTGGATCGGCTCACAGACTCAGAGCGTGAAGTGAATGG 1500
|||||
Db 1683 ATGGTAGGAGAGAGAGAGCGCACTGGATCGGCTCACAGACTCAGAGCGTGAAGTGAATGG 1742
|||||
```

```
QY 1501 AAGTGGCTGATGGGACATCTCCAGACTACAAAAATTGGAAGCTGGACAGCCGGATAAC 1560
|||||
Db 1743 AAGTGGCTGATGGGACATCTCCAGACTACAAAAATTGGAAGCTGGACAGCCGGATAAC 1802
|||||
QY 1561 TGGGTCATGGCCATGGCCAGGAGAGACTGCTGGGTTGATTATTATGCTGGCAGTGG 1620
|||||
Db 1803 TGGGTCATGGCCATGGCCAGGAGAGACTGCTGGGTTGATTATTATGCTGGCAGTGG 1862
|||||
QY 1621 AAGGATTTCCAATGTGAAGACGCTCAATAACTTTCATTTGCGAAAAAGACAGGGAGACAGTA 1680
|||||
Db 1863 AAGGATTTCCAATGTGAAGACGCTCAATAACTTTCATTTGCGAAAAAGACAGGGAGACAGTA 1922
|||||
QY 1681 CTGTCATCTGCATTA 1695
|||||
Db 1923 CTGTCATCTGCATTA 1937
|||||

RESULT 3
AB038518      3058 bp mRNA linear PRI 08-MAR-2001
LOCUS      Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
DEFINITION      type I, complete cds.
ACCESSION      AB038518
VERSION      AB038518.1 GI:13365514
KEYWORDS
SOURCE      Homo sapiens tissue_lib:Placenta cDNA to mRNA.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Nakamura,K., Funakoshi,H., Miyamoto,K., Tokunaga,F. and Nakamura,T.
Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRCL), a novel member of a
scavenger receptor family
Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
2 (bases 1 to 3058)
Nakamura,K. and Nakamura,T.
Direct Submission
Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical
School, Division of Biochemistry, Biomedical Research Center, 2-2
Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail:knakamura@onbich.med.osaka-u.ac.jp,
Tel:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)
Location/Qualifiers
1..3058
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_lib="Placenta"
1..3058
/gene="SRCL"
209..2437
/gene="SRCL"
/codon_start=1
/product="scavenger receptor with C-type lectin type I"
/protein_id="BAB39147.1"
/db_xref="GI:13365515"
/translation="MKDDFAEEEEEVSFGYKRFGIQEQTCFKKNWALKFESIILY
ILCALLTIVAILGKVKVMDNVNTGGMTESRQTVDDKLTAVESDLKLKLDGDTGKAI
STNSLSTFRSDILDRLROOLREITEKTSKNKDTLEKLOASGDALVDRQSOLKETLENN
SFLLITVNTKLAQYNGVYTNLQODTSVLQGNLQNMYSHVIMNMLNLTQVQQRN
LITNQRSVDDTSQAIRIKNDQVFNQQLQVFLQAKKDDTLKDKVQSLQTLAANNLSALA
KANNDTLEDMSQLSFTQGMENITTSIQANEQNLDQLDKHKAENRTAIFNQLAE
RPLQFETDVTIISNYSYTAHLRLTSLNLEVRTCTDTLTKHDDLTSLNLTLANI
RUDSVLRMOODLSRLDTEVANLSVTMEEMKLVDSKHGOLIKNFTILOQPPGPRGP
RGDRSQGPPGPTGNKGOKGKEGEPGPGPAGERPIGPAGPPGEGKSGKSGOPK
SGRSPGPPGPGSGDPPGPPGKEGLPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
LPVPMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
KEIFEDAKLFCDSKSHLVFINTRREQQWIKQKQMGVRESHWIGLTDLSRENENKWLID
TSPDYKNWKAQPDNKGHGHGPGEDCAGLIYAGQWNDFQEDVNNFICEKDRETVLSS
AL"
polya_signal      3020..3025
/gene="SRCL"
```

```
BASE COUNT      911 a      744 c      764 g      639 t
ORIGIN
Query Match      99.8%; Score 1691.8; DB 9; Length 3058;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1693; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GTCAACAATCTGCAGCAAGATACACAGCTGCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
|||||
Db 740 GTCAACAATCTGCAGCAAGATACACAGCTGCTCCAGGGCAATCTGCAGAACCAAAATGTAT 799
|||||
QY* 61 TCTCATAATGTGTGTATCATGAACCTCAACAACCTGAACTGACCCAGGTGTCAGCAGAGG 120
|||||
Db 800 TCTCATAATGTGTGTATCATGAACCTCAACAACCTGAACTGACCCAGGTGTCAGCAGAGG 859
|||||
QY 121 AACCTCATCAGGAATCTGCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 180
|||||
Db 860 AACCTCATCAGGAATCTGCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 919
|||||
QY 181 AAGAAGCACTTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGAGGACACGGATTGG 240
|||||
Db 920 AAGAAGCACTTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGAGGACACGGATTGG 979
|||||
QY 241 CTGAAGAGAGAAAGTCAGAGCTTGCAGACGCTGGCTGCCAACAACCTCTCGGTTGGCCAAA 300
|||||
Db 980 CTGAAGAGAGAAAGTCAGAGCTTGCAGACACTGGCTGCCAACAACCTCTCGGTTGGCCAAA 1039
|||||
QY 301 GCCAACAAGCACACCCTGGAGGATATGAACAGCCAGCTCAACTCATTTCACAGGTCAGATG 360
|||||
Db 1040 GCCAACAAGCACACCCTGGAGGATATGAACAGCCAGCTCAACTCATTTCACAGGTCAGATG 1099
|||||
QY 361 GAGAACATCACCCTACTCTCTCAAGCCAAAGCAGCAGCAACCTGAAAGACCTGTCAGGACTTA 420
|||||
Db 1100 GAGAACATCACCCTACTCTCTCAAGCCAAAGCAGCAGCAACCTGAAAGACCTGTCAGGACTTA 1159
|||||
QY 421 CACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCACTGGAGGAACCTTCCAG 480
|||||
Db 1160 CACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCACTGGAGGAACCTTCCAG 1219
|||||
QY 481 CTCTTTGAGACGGATATTGTGAACATCATTAGCAATATATAGTTTACACAGCCACCACTG 540
|||||
Db 1220 CTCTTTGAGACGGATATTGTGAACATCATTAGCAATATATAGTTTACACAGCCACCACTG 1279
|||||
QY 541 CGGACGCTGACCAGCAATCTAAATGAAGTCAGACCACTTGCACAGATACCCCTTACCAAAA 600
|||||
Db 1280 CGGACGCTGACCAGCAATCTAAATGAAGTCAGACCACTTGCACAGATACCCCTTACCAAAA 1339
|||||
QY 601 CACACAGATGATCTGACCTCTTGAATATATACCTGGCCAAACATCCGTTTGGATTCTGTT 660
|||||
Db 1340 CACACAGATGATCTGACCTCTTGAATATATACCTGGCCAAACATCCGTTTGGATTCTGTT 1399
|||||
QY 661 TCTCTCAGGATGCAACAAGATTTGATGAGGTCGAGGTTAGACACTGAAGTAGGCAACTTA 720
|||||
Db 1400 TCTCTCAGGATGCAACAAGATTTGATGAGGTCGAGGTTAGACACTGAAGTAGGCAACTTA 1459
|||||
QY 721 TCAGTGATTATGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAAT 780
|||||
Db 1460 TCAGTGATTATGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAAT 1519
|||||
QY 781 TTTTACAATACTACAAGGTCACCGGGCCCCAGGGTCCAGAGGTGACAGAGATCCCAG 840
|||||
Db 1520 TTTTACAATACTACAAGGTCACCGGGCCCCAGGGTCCAGAGGTGACAGAGATCCCAG 1579
|||||
QY 841 GGACCCCTTGGCCCAACTTGCAACAAGGGACACAAAAGGAGAGAGAGGGAGCCCTGGACCA 900
|||||
Db 1580 GGACCCCTTGGCCCAACTTGCAACAAGGGACACAAAAGGAGAGAGAGGGAGCCCTGGACCA 1639
|||||
QY 901 CTGGCCCTTGGCCCTGAGAGAGGCCCAATTTGGACCAAGCTGGTCCCCCGGAGAGCGTGGC 960
|||||
Db 1640 CTGGCCCTTGGCCCTGAGAGAGGCCCAATTTGGACCAAGCTGGTCCCCCGGAGAGCGTGGC 1699
|||||
QY 961 GGCAAAGGATCTAAAGGCTCCCAAGGGCCCCCAAGGCTCCGTTGTTCCCCCTGGGAAGCCC 1020
|||||
```


| | | | |
|----------------------------|--|--|----------------------------|
| Db | 848 | GGGACCCAGGCCCCCGGGGCCCCACAGGACCAAGAGGACACTCCCGGGCCCTCAGGGCCCTC | 907 |
| Qy | 1100 | CTGGCTTCCAGGAGACTTCAGGCGCACCGCTTGGGAGCCCTGGGGTGCCTGGACCTCGGGAC | 1159 |
| Db | 908 | CTGGCTTCCAGGAGACTTCAGGCGCACCGCTTGGGAGCCCTGGGGTGCCTGGACCTCGGGAC | 967 |
| Qy | 1160 | TGCCAGGCTTGCCTGGGGTACCAAGCATGCGACGCCCAAGGGCCCCCGGCCCTCCTG | 1219 |
| Db | 968 | TGCCAGGCTTGCCTGGGGTACCAAGCATGCGACGCCCAAGGGCCCCCGGCCCTCCTG | 1027 |
| Qy | 1220 | GCCCATCAGGAGCGGTGGTGCCTTGGCCCTGCGAGATGAGCCAAACCCGCGACCGGAGG | 1279 |
| Db | 1028 | GCCCATCAGGAGCGGTGGTGCCTTGGCCCTGCGAGATGAGCCAAACCCGCGACCGGAGG | 1087 |
| Qy | 1280 | ACAATGGCTGCCGCCCTCACTTGAAGAAGCTTCACAGACAAATGCTACTATTTTCAGTTG | 1339 |
| Db | 1088 | ACAATGGCTGCCGCCCTCACTTGAAGAAGCTTCACAGACAAATGCTACTATTTTCAGTTG | 1147 |
| Qy | 1340 | AGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTT | 1399 |
| Db | 1148 | AGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTT | 1207 |
| Qy | 1400 | TCATAAACACATAGAGAGGACAGCAAGTGGATATAAAAAACAGATGGTAGGAGAGAGGCC | 1459 |
| Db | 1208 | TCATAAACACATAGAGAGGACAGCAAGTGGATATAAAAAACAGATGGTAGGAGAGAGGCC | 1267 |
| Qy | 1460 | ACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGAAGTGGCTGGATGGGACAT | 1519 |
| Db | 1268 | ACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGAAGTGGCTGGATGGGACAT | 1327 |
| Qy | 1520 | CTCAGAGACTACAAAAATTGGAAGCTGGACGCCGATTAACCTGGGGTCAATGGCCATGGGC | 1579 |
| Db | 1328 | CTCAGAGACTACAAAAATTGGAAGCTGGACGCCGATTAACCTGGGGTCAATGGCCATGGGC | 1387 |
| Qy | 1580 | CAGGAGAGACTGCTGGGGTGTATTTATGCTGGCAGTGAACGATTTCCAATGTGAAG | 1639 |
| Db | 1388 | CAGGAGAGACTGCTGGGGTGTATTTATGCTGGCAGTGAACGATTTCCAATGTGAAG | 1447 |
| Qy | 1640 | ACGTCAATAACTTCATTTGCCAAAAAGACAGGAGACAGTACTGTCATCTGCATTA | 1695 |
| Db | 1448 | ACGTCAATAACTTCATTTGCCAAAAAGACAGGAGACAGTACTGTCATCTGCATTA | 1503 |
| RESULT 5 | | | |
| LOCUS | AX490920 | 2005 bp | DNA linear PAT 16-AUG-2002 |
| DEFINITION | Sequence 27 from Patent WO0200690. | | |
| ACCESSION | AX490920 | | |
| VERSION | AX490920.1 | GI:22323797 | |
| KEYWORDS | human. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo. | | |
| AUTHORS | Baker, K.P., Ferrata, N., Gerber, H., Gerritsen, M.E., Goddard, A., Godowski, P.J., Gurney, A.L., Hillan, K.J., Marsters, S.A., Pan, J., Paoni, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I. and Ye, W. | | |
| TITLE | Compositions and methods for the diagnosis and treatment of disorders involving angiogenesis | | |
| JOURNAL | Patent: WO 0200690-A 27 03-JAN-2002; | | |
| FEATURES | Genentech, Inc. (US) | | |
| source | Location/Qualifiers | | |
| | 1..2005 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| BASE COUNT | 606 a | 493 c | 491 g 415 t |
| ORIGIN | | | |
| Query Match | 87.2% | Score 1477.6; | DB 6; Length 2005; |
| Best Local Similarity | 99.7% | Pred. No. 0; | |
| Matches 1491: Conservative | 0; | Mismatches | 4; Indels 1; Gaps 1; |

| | | | |
|------------|--|---|-----------------------------|
| Db | 1463 | GGACCACCTGCTGCTCAACTGGCAACAAGGGACAGAAAAGGAGAGAGAGCGCTGGTGCCA | 1522 |
| Qy | 901 | CCTGGCCCTCGGGGTGAGAGAGAGCCCAATTGGACACAGCTGGTCCCCCGGAGAGCGGTGGC | 960 |
| Db | 1523 | CCTGGCCCTCGGGGTGAGAGAGAGCGACAATTGGACACAGCTGGCCCTCTCTGGAGAGCGGTGGC | 1582 |
| Qy | 961 | GGCAAGAGATCTAAAGGCTCCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCCCTGGGAAAGCCC | 1020 |
| Db | 1583 | AGCAAGAGATCCAAAGGCTCACAGGGTCCCAAGAGATCTCTGGTGGTCCCCCAGGGAAGCCT | 1642 |
| Qy | 1021 | GGCCCTCAGGGCCCCAGTGGGGACCCAGGCCCCCGGGGCCACACAGGCAAAAGAGGGACTC | 1080 |
| Db | 1643 | GGCCCTCAAGGACCTAGTGGGGACCCAGGACCAACAGTCCACAGGCCAAGATGGACTC | 1702 |
| Qy | 1081 | CCGGCCCTCAGGGCCCTCTCGGCTTCAGGGACTTCAGGGACCTTCAGGGACCGTGGGGAGCCCTGGG | 1140 |
| Db | 1703 | CCTGGCCCTCAGGGCCCTCTCGGCTTCAGGGACTACAGGGACTGTGGGTGAGCCCTGGGA | 1762 |
| Qy | 1141 | GTGCTTGGAACCTCGGGGACTGCCAGGCTTGCTGGGGTACCAGGCATGCCAGGCCCTTAAG | 1200 |
| Db | 1763 | GTACCTGACCTCGGGGTTGCCAGGCTGCCAGGGTGCACAGGCATGCCCTGGGCCCTAAG | 1822 |
| Qy | 1201 | GGCCCCCGGGCCCTCTCTGGCCCATCAGGAGCGGTGTGGCCCTGGGCCCTGCAGAAATGAG | 1260 |
| Db | 1823 | GGACACGTGGCCCTCCAGGGCCCTCAGGAGCAATGGAGCCATTGGCTCTGCAGAAATGAA | 1882 |
| Qy | 1261 | CCAACCCGGCACGAGGACAATGGCTGCCGCCCTCAGTGGAGAACTTCACAGACAAA | 1320 |
| Db | 1883 | CCAACCCAGCATCAGAGGTACAGCGATGTCGGCTCAGTGGAGAACTTCACAGATAAA | 1942 |
| Qy | 1321 | TGCTACTATTTTTTCAGTTGAGAAAAGAAATTTTTGAGGATGCAAAAGCTTTCTGTGAAGAC | 1380 |
| Db | 1943 | TGCTACTATTTTTTCATTTGGAAGAAAGAAATTTTCAAGATGCTAAGCTTTCTGTGAAGAC | 2002 |
| Qy | 1381 | AAGTCTTCACATCTGTTTTCATAAAGACATAGAGAGAACAGCAATGGATAAAAAACAG | 1440 |
| Db | 2003 | AAATCTTCCCATCTCGTTTTTCATAAACTCAAGAGAAGAACAGCAATGGATAAAAAAGCAT | 2062 |
| Qy | 1441 | ATGTTAGGGAGAGAGCCACTGGATCGGCTCACAGCTCAGAGCGTGAAAATGAATGG | 1500 |
| Db | 2063 | ACCTGGGGAGAGAAAGCCATTGGATCGGCTCACAGCTCAGAAAGAGAAAGCAATGG | 2122 |
| Qy | 1501 | AAGTGGCTGGATGGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCCGGATAAC | 1560 |
| Db | 2123 | AAGTGGCTAGAGGGTCACCTGTTGATTACAAAAGCTGGAAAGCTGGACACACAGATAAC | 2182 |
| Qy | 1561 | TGGGGTCATGGCCATGGCCAGAGAGACTGTGCTGGGTGATTATGCTGGGCGAGTGG | 1620 |
| Db | 2183 | TGGGGCAGTGGCCATGGCCAGGAGAGACTGTGCTGGCTTGATTACGCGAGGACAGTGG | 2242 |
| Qy | 1621 | AACGATTTCGAATGTGAAGAGTCAATAACTTCATTTCCGAAAAGACAGGGAGACAGTA | 1680 |
| Db | 2243 | ATGACTTCCAGTGTGATGAATCAATAACTTCATTTGTGAGAAGAAAGGGAGGCGAGTA | 2302 |
| Qy | 1681 | CTGTCTCTGTCATTA | 1695 |
| Db | 2303 | CCATCATCCATATTA | 2317 |
| RESULT 7 | | | |
| AB038519 | | | |
| LOCUS | AB038519 | 3291 bp | mRNA linear ROD 18-JAN-2002 |
| DEFINITION | Mus musculus scrl mRNA for scavenger receptor with C-type lectin, | | |
| ACCESSION | complete cds. | | |
| VERSION | AB038519 | | |
| KEYWORDS | AB038519.1 GI:18146951 | | |
| SOURCE | Mus musculus cDNA to mRNA. | | |
| ORGANISM | Mus musculus | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| | 1 Nakamura, K., Funakoshi, H., Tokunaga, F. and Nakamura, T. | | |

TITLE Molecular cloning of a mouse scavenger receptor with C-type lectin (SRC1)(1), a novel member of the scavenger receptor family
JOURNAL Biochim. Biophys. Acta 1522 (1), 53-58 (2001)
MEDLINE 21575692
REFERENCE 2 (bases 1 to 3291)
AUTHORS Nakamura,K. and Nakamura,T.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical School, Division of Biochemistry, Biomedical Research Center; 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
FEATURES (E-mail:knakamur@onbich.med.osaka-u.ac.jp, Tel:81-6-6879-3783(Fax:3783), Fax:81-6-6879-3789)
source Location/Qualifiers
1..3291
/organism="Mus musculus"
/db_xref="taxon:10090"
gene 1..3291
/gene="src1"
CDS 77..2305
/gene="src1"
/codon_start=1
/product="scavenger receptor with C-type lectin"
/protein_id="BAB82497.1"
/db_xref="GI:18146952"
/translation="MKDDFAEEVEVSFGYKRFGIHEGTQCKINNALKFSIVLL
ILCALITVAIIIGYKVKEMDNVSDGMSHOTYDNKLTAVESDLKLGDOAGKKAL
STNSLSTFRSDILDRLROOLOETEXTSKNKTDLKLAQGDLSVDROSLETQGN
SFLLTVNKTQAYNGYVNLQDYNVLOGNLQSQMSQSVYIMNLNLNLTVQVQRN
LISNLQSDVDTLSA IQRKNQFQVFLQAKDQTDMLKVKVQSLQTLAAANSLA
KANNDLIDNMSGLSFGTGMNDITTSIQANEQSLQDLQDKHKTENRTAVKPSLEE
RFQVETDIWIIISNTYSYTAHLRTLSNLNDVTTCTDTLTHTDDTLTSLNLTVNI
RLOS ISLRMDMEEMRSKLDTEVANSLSVVMMEKLVDSKHQLIKNFTILQGPGRPG
KDRSGQPGPTGNGKQGEKGPVGPAGEGTIGVPVGPGRGSGKSGSGQPK
LGRSGPGKPGQSPGDPGPPGPPGKDPGLPGPPGFLQGTGVEGVPGRGLPG
LSPVMPGPGKPGSGTGMGMEPLAQNEPTPASEVNGCPPHWKFNFTDKCYFYLE
KETLEKALFCDEKSSHLVFINSRREQWIKKHTVGRSHWIGLTDSQEESEMKWLDG
SPVDYKKNRWAGQPDNNGSGHGPGECDAGLIYAGQWDFQCEIDNNFICBKREAVPSS
IL"
polyA_signal 3246..3251
/gene="src1"
BASE COUNT 1013 a 757 c 759 g 762 t
ORIGIN
Query Match 79.9%; Score 1354.2; DB 10; Length 3291;
Best Local Similarity 87.4%; Pred. No. 1.3e-308;
Matches 1482; Conservative 0; Mismatches 213; Indels 0; Gaps 0;
Qy 1 GTCACGAATCTGCAGCAAGATACCAAGCTGCTCCAGGCGCAATCTGCAGAACCAATGAT 60
Db 608 GTCACAAATCTGCACAAAGATACTAATCTGCTCCAGGCGCAATCTGCAGAGCAATGAT 667
Qy 61 TCTCATATGTGTCATCATGAACCTCAACAACTGAACCTGACCCAGGTGCACACAGG 120
Db 668 TCTCAGACGCTGGTATTATCATGAAGCTCAACAACTGAACCTAACCCAGGTTCAGACAGG 727
Qy 121 AACCTCATCAGCAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGAATC 180
Db 728 AACCTTATCTCAAAATCTGCAGCAGTCTGTGGATGACACAAAGCCGTGGCCATCCAGCAATT 787
Qy 181 AAGAACGACTTTCAAAATCTGCAGCAGGTTTTCTTCAAGCCAAAGAGGACACAGGATGG 240
Db 788 AAGAATGATTTCACAAATCTGCAGCAGGTTTTCTTCAAGCCAAAGAGGACACCGATTGG 847
Qy 241 CTGAAGAGAAAGTGCAGAGCTTCGACACGCTGCGTCCCAACACTCTCGCTGGTGGCCAAA 300
Db 848 CTAAGAGAAAAGTGCAGAGCTTCGACACATFTGCTCTCCCAACACTCTGCGCTGGCCAAA 907
Qy 301 GCCAACAGGACACCCCTGGAGGATATGAACAGGCGAGTCAACTCATTTACAGGTCAGATG 360
Db 908 GCCAACATGACACCCCTAGAGGATATGAATAGCCAGCTCAGCTCATTTACAGGTCAGATG 967
Qy 361 GAGAACATCACCACTATCTCTCAAGCCCAAGCAGCAACCTGAAAGACCTTCGAGGACTTA 420
Db 420 GAGAACATCACCACTATCTCTCAAGCCCAAGCAGCAACCTGAAAGACCTTCGAGGACTTA

Db 968 GACACATTACACATATCTTCACAGGCCAACAGAGAGAGCCTTGAAGACCTTCAGGACTTA 1027
QY 421 CACAAAGATGCAGAAATAGAACAGCCATCAAGTTCAACCACTGAGGAAGCGCTTCAG 480
Db 1028 CACAAAGATACAGAAATAGAACAGCTGTCAAGTTACGCCAACTTGAAGAACGCTTCCAG 1087
QY 481 CTCCTTTGAGACGGATATTTGTGAACATCATATTGCAATATCAGTTTACAGACCCACACCTG 540
Db 1088 GTCTTTGAGACAGATATTTGTGAACATCATATTGCAACATCAGCTACACAGCCCATCACCTG 1147
QY 541 CGGACGCTGACCAGCAATCTTAATGAAGTTCAGGACCCACTTGCACAGATACCTTACCAAAA 600
Db 1148 AGACACTGACACCAACATCTGAATGATGTTGGACCAATGACACAGACACCTTGACACAGA 1207
QY 601 CACACAGATGATCTGACCTCTTGAATATACCTTGGCCAAACATCCGTTTGATTTCTGTT 660
Db 1208 CACACGGATGACCTGACCTCTTGAATTAACACACTAGTCAACATCCGCTGGATTTCTATT 1267
QY 661 TCTCTCAGGATGCAACAGATTTTGATGAGTTCGAGCTTAGACACTGAAGTAGCCCAACTTA 720
Db 1268 TCTCTCAGGATGCAACAGATGATGAGTTCAGGTTCAAAAGTTAGACACTGAAGTGCCCAACTTA 1327
QY 721 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGCTCAGCTCATCAAGAAAT 780
Db 1328 TCAGTGATTATGGAAGAGATGAAGCTGTTGACTCCAAGCAGCGTCAAGCTCATCAAGAAC 1387
QY 781 TTTACAATACTACAAGGTCCACCGGCCCCCGAGGCTGCTGACAGAGTGCAGAGATCCCGAG 840
Db 1388 TTTACCAATTTACAAGGTCTCTGCTGCCCCAGAGGTCCAAGAGGTGACAGAGGATCTCAG 1447
QY 841 GGACCCCTTGGCCCAACTGGCAAGGAGGACAGAAAGGAGAGAGAGGGGAGCCTGGACCA 900
Db 1448 GGACCACTGTGTCACACTGGCAAGGAGGACAGAAAGGAGAGAGAGGAGCCTGGTCCA 1507
QY 901 CCTGCGCTTGGCGGTGAGAGAGGCCAATTTGACACAGCTGCTGCCCGGAGAGCGTGGC 960
Db 1508 CTTGCGCTTGGCGGTGAGAGGGGACAAATTTGACACAGCTGCTGCCCGGAGAGCGTGGC 1567
QY 961 GCGAAAGGATCTAAAGGCTCCAGGGCCCCCAAGGCTCCCGTGGTTCCTCGGGAAGCCC 1020
Db 1568 AGCAAGGATCCAAAGGCTCACAGGTCCTCAAGAGTCTCGTGGTTCCTCGGGAAGGCT 1627
QY 1021 GCGCTCAGGGCCCCAGTGGGACCCAGGCCCGGCCCCCAGGAGGAGGAGGAGGAGTCC 1080
Db 1628 GCGCTCAGGAGACCTAGTGGGACCCAGGACCCAGGCTCCAGGAGGAGGAGGAGTCC 1687
QY 1081 CCGGCGCTCAGGGCCCTCCTGGCTTCCAGGAGCTTCAGGGACCCGTTGGGAGCGCTGGG 1140
Db 1688 CTTGCGCTCAGGGCCCTCCTGGCTTCCAGGAGCTTCAGGGAGCTGTTGGTGGAGCTGGA 1747
QY 1141 GTGCTTGAGCTCGGGGACTGCCAGCTTGCTGGGGTACAGGAGTCCAGGCCCCCAAG 1200
Db 1748 GTACCTGAGCTCGGGGGTTCAGGCTTGCCAGGGGTGCGAGGATGCTTGGGCTTAAG 1807
QY 1201 GCGCCCCCGGCTTCCTGGGCCATCAGAGCGGTGTTGGCCCTCGGCTCGAGAAATGAG 1260
Db 1808 GGAACACCTTGGCCCTTCAGGCCCCCTCAGGAGCAATGGAGCCATTTGGCTTGAGAAATGAA 1867
QY 1261 CCAACCCCGGACCCGAGGACAAATGGCTGCCCGCTCACTGGAAGAACTTCAACAGACAAA 1320
Db 1868 CCAACCCCGAGCATCAGAGTCAAGGATGTCGCGCTCACTGGAAGAACTTCAACAGATAA 1927
QY 1321 TGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCTTTCTGTGGAAGAC 1380
Db 1928 TGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGATGCTTAAGCTTTCTGTGGAAGAC 1987
QY 1381 AAGTCTTCACATCTTTTTCATAAACAACACTAGAGAGAACAGCAATGAGNATAAAAAACAG 1440
Db 1988 AAATCTTCCCATCTGTTTTCATAAACAACACTAGAGAGAACAGCAATGAGNATAAAAAACAT 2047
QY 1441 ATGGTAGGAGAGAGAGCCACTGGATCGGCCTCAGAGCTCAGAGCGTGAAGAAATGAATGG 1500
Db 2048 ACCGTGGGAGAGAAAGCCATTGGATCGGCCTCAGAGCTCAGAGAGAAAGCGAATGG 2107

QY 1501 AAGTGGCTGGATGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGATAC 1560.
Db 2108 AAGTGGCTAGAGCGGTACCTGTGATTACAAAAACTGGAAAGCTGGACAAACAGATAC 2167
QY 1561 TGGGTCATGTCATGGCCATGGCCAGGAGAACTGTGCTGGTTGATTATCTGCTGGCAGTGG 1620
Db 2168 TGGGTCATGTCATGGCCATGGCCAGGAGAACTGTGCTGGTTGATTATCTGCTGGCAGTGG 2227
QY 1621 AACGATTTCCATGTTGAAGACGTCATTAATCACTTCTTTGGGAAAAAGACAGGAGACAGTA 1680
Db 2228 AATGATTTCCATGTTGAAGATCAATCACTTCTTTGGGAAAAAGGAGGAGGAGTA 2287
QY 1681 CTGTCATCTGCATTA 1695
Db 2288 CCATCATCATATTA 2302

RESULT 8
AB052103
LOCUS
DEFINITION Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
type II, complete cds.
ACCESSION AB052103
VERSION AB052103.1 GI:13365552
KEYWORDS
SOURCE Homo sapiens cdna to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (sites)
Nakamura,K., Funakoshi,H., Miyamoto,K., Tokunaga,F. and Nakamura,T.
Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRCL), a novel member of a
scavenger receptor family
Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
21092718
2 (bases 1 to 4330)
Nakamura,K. and Nakamura,T.
Direct Submission
Submitted (04-DEC-2000) Kenji Nakamura, Osaka University Graduate
School of Medicine, Division of Biochemistry, Biomedical Research
Center; 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail:knakamur@onbich.med.osaka-u.ac.jp, Tel:81-6-6879-3783,
Fax:81-6-6879-3789)

FEATURES
Location/Qualifiers
1. 4330
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. 4330
/gene="SRCL"
3. 1871
/gene="SRCL"
/codon_start=1
/product="scavenger receptor with C-type lectin type II"
/protein_id="BAB39148.1"
/db_xref="GI:13365553"
/translation="MKDDFAEEVEVPSFGFKRFGHEGTQCPKCHNNWAKFSILLY
ILCALLTITVAIILGYKVVEKMDNTGGMETSRQVDDKLTAVESDLKILGDOTGKAI
STNSELSTFRSDILDRLQQLREITEKTSKNDTLEKLAQSGDALVDROSQLETLNN
SFLITTVNKTLOAYVYNNLQQDTSLVQLNOMYSHNVVYNNLNLTVQVORN
LITNLRSDVDTQAIRIKNDQVFLQAKKDTDLKLEKVKQSLQTLAANSALA
KANNDTLEDNMSQLNSFTGMENTTTISQANEQNLKLDLHKDAENRTAKFNQLEE
RFOLFETDVNIISNYSVAHLRTLSNLNEVRTCTDTLTHTDDTLTNNLTANI
RLDSVLRMOQDLMRSLDTEVNLVMEEMKLVDSKHQOLIKFTTLOGPGRGP
RGDRSGSPGPTNGKQKGEKGPFGPAGERGPIGPAGPGRGKSGKSGQPK
GSRGSPKPGPGPGPPGPGKGLPGPPGPGGLOGTVEGVPVPGKGLPG
LPVPGMPGPKGPGPGPGPVPLALQNEPTPAEDNSKSPSLQPGGGSACA"
BASE COUNT 1186 a 1065 c 1035 g 1044 t
ORIGIN
Query Match 75.6%; Score 1281.2; DB 9; Length 4330;
Best Local Similarity 99.8%; Pred. No. 2.2e-291;

Matches 1283; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 GTCAAGAACTCTGAGCAAGATACACGGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
Db 534 GTCAAGAACTCTGAGCAAGATACACGGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 593
QY 61 TCTCATAAATGTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGACAGAGG 120
Db 594 TCTCATAAATGTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGACAGAGG 653
QY 121 AACTCATACGAATCTGCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 180
Db 654 AACTCATACGAATCTGCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 713
QY 181 AAGAAGCACTTTCAAAATCTGCAGAGGTTTTTCTCAAGCCCAAGAGGACCGAATGG 240
Db 714 AAGAAGCACTTTCAAAATCTGCAGAGGTTTTTCTCAAGCCCAAGAGGACCGAATGG 773
QY 241 CTGAAGGAGAAAGTCGAGAGCTTTCGAGAGCTGGCTGCCAACAACCTCTGCTTGGCCAAA 300
Db 774 CTGAAGGAGAAAGTCGAGAGCTTTCGAGAGCTGGCTGCCAACAACCTCTGCTTGGCCAAA 833
QY 301 GCCAACAAGCAGACACCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGTCAAGT 360
Db 834 GCCAACAAGCAGACACCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGTCAAGT 893
QY 361 GAGAACATCACCACTATCTCTCAAGCCCAAGCAGCAGCAACCTGAAAGACCTGCAGGACTTA 420
Db 894 GAGAACATCACCACTATCTCTCAAGCCCAAGCAGCAGCAACCTGAAAGACCTGCAGGACTTA 953
QY 421 CACAAGATGCAGAGAAATGAGACGCGATCAAGTTCAACCAACTGGAGGAACCTTCCAG 480
Db 954 CACAAGATGCAGAGAAATGAGACGCGATCAAGTTCAACCAACTGGAGGAACCTTCCAG 1013
QY 481 CTCTTGTAGACGGATTTGTGACATCATTAGCAATATCAGTTACACAGCCACACCTG 540
Db 1014 CTCTTGTAGACGGATTTGTGACATCATTAGCAATATCAGTTACACAGCCACACCTG 1073
QY 541 CGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACACCTTGACAGATACCCCTTACCAAA 600
Db 1074 CGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACACCTTGACAGATACCCCTTACCAAA 1133
QY 601 CACACAGATGATCTGACCTCTTGAATAATACCTGGCCAAACATCCGTTGGATTCTGTT 660
Db 1134 CACACAGATGATCTGACCTCTTGAATAATACCTGGCCAAACATCCGTTGGATTCTGTT 1193
QY 661 TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTAGACACTGAGTACAGTACCAACTTA 720
Db 1194 TCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTAGACACTGAGTACAGTACCAACTTA 1253
QY 721 TCAGTGATTATGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAAT 780
Db 1254 TCAGTGATTATGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAAT 1313
QY 781 TTTTACAATACTACAAGTCCACCGGGCCCCAGGGTCCAGAGGTACAGAGATCCCAAG 840
Db 1314 TTTTACAATACTACAAGTCCACCGGGCCCCAGGGTCCAGAGGTACAGAGATCCCAAG 1373
QY 841 GGACCCCTTGGCCCAACTTGGCAACAAGGGACAGAAAGAGGAGAAAGGGAGCCTGGACCA 900
Db 1374 GGACCCCTTGGCCCAACTTGGCAACAAGGGACAGAAAGAGGAGAAAGGGAGCCTGGACCA 1433
QY 901 CTTGGCCCTTGGCCCTTCCAGGCCCCCAAGGCTCCCGTGGTTCCTTCCCTGGGAAGCCC 960
Db 1434 CTTGGCCCTTGGCCCTTCCAGGCCCCCAAGGCTCCCGTGGTTCCTTCCCTGGGAAGCCC 1493
QY 961 GGCAAGGATCTAAAGCTTCCAGGCCCCCAAGGCTCCCGTGGTTCCTTCCCTGGGAAGCCC 1020
Db 1494 GGCAAGGATCTAAAGCTTCCAGGCCCCCAAGGCTCCCGTGGTTCCTTCCCTGGGAAGCCC 1553
QY 1021 GGCCCTCAGGGCCCCAGTGGGAGCCAGGCCCCCGGGCCCCACAGGCAAGAGGAGCTC 1080
Db 1554 GGCCCTCAGGGCCCCAGTGGGAGCCAGGCCCCCGGGCCCCACAGGCAAGAGGAGCTC 1613
```

```
QY 1081 CCGGCCCTCAGGCGCTCTCTGCTTCCAGGGACTTTCAGGGCACCGTTGGGGAGCCTGGG 1140
Db 1614 CCGGCCCTCAGGCGCTCTCTGCTTCCAGGGACTTTCAGGGCACCGTTGGGGAGCCTGGG 1673
QY 1141 GTGCTTGGACCTCGGGGACTTGCAGGCTTGCCTGGGTACCAAGGCATGCCAGCCCAAG 1200
Db 1674 GTGCTTGGACCTCGGGGACTTGCAGGCTTGCCTGGGTACCAAGGCATGCCAGCCCAAG 1733
QY 1201 GGGCCCCCGGCGCTCTCTGCGCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAG 1260
Db 1734 GGGCCCCCGGCGCTCTCTGCGCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAG 1793
QY 1261 CCAACCCCGGCACCGGAGGACAATGG 1286
Db 1794 CCAACCCCGGCACCGGAGGACAATAG 1819
```

```
RESULT 9
HSM802985
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
HSM802985 1886 bp mRNA linear PRI 20-MAR-2002
Homo sapiens mRNA; cDNA DKF2p547G1215 (from clone DKF2p547G1215).
AL713657
AL713657.1 GI:19584339
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Koehler, K., Beyer, A., Mewes, H.W., Weil, B. and Wiemann, S.
Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKF2p547G1215) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
```

```
FEATURES
Location/Qualifiers
1..1886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18p11.3"
/clone="DKF2p547G1215"
/tissue_type="brain"
/clone_lib="547 (synonym: hfbr1). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="fetal"
1..1886
/gene="DKF2p547G1215"
<1..1256
/gene="DKF2p547G1215"
/note="similarity to asialoglycoprotein receptor (Mus
musculus"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAD28466.1"
/db_xref="GI:19584340"
/translation="AIKFNLEERFQLPETDIVNIISNYSYFAHLRLTSLNLEVRT
TCTDPLTKHTDOLTSLNLTANIRLDSVSLRMOOLMBSRLDTEVANLSVIMEKLV
DSKHQLIKNFITLQPGPRGDRSGQPGPTGNKQKGEKGEKGGPPGAGERG
PIGPAGPEERGGKSGQSGPKSGRSGPGSGSDGPPGPPGPPGKGLPGQGP
PFGQLQGTGVEGVPYGPGLPGVFGMPGKPGPGPGPGVAVPLALQNEPTPA
PDNCGPPHKNFTDKCYFYFVSEKEIFEDAKLFCEDKSKSLVFINTRREQQWIKOMV
GRESHWIGLTDSERENKWLDTGTPDYKNKAGOPDNWNGHGGPGEDCAGLIYAGQW
NDFOCEDVNNFICEKRETVLSSAL"
1838..1843
polya_signal
```

```
/gene="DKF2p547G1215"
1869
polyA_site
/gene="DKF2p547G1215"
BASE COUNT 569 a 444 c 445 g 428 t
ORIGIN

Query Match 73.8%; Score 1251.4; DB 9; Length 1886;
Best Local Similarity 99.9%; Pred. No. 2.1e-284;
Matches 1252; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 443 CAGCCATCAAGTTCAACCACTGGAGAACGCTTCACGCTCTTGAGCGGATATTGTCA 502
DB 1 CAGCCATCAAGTTCAACCACTGGAGAACGCTTCACGCTCTTGAGCGGATATTGTCA 60
QY 503 ACATCATTAAGCAATATCAGTTACACAGCCCAACACCTCGCGAGCGCTGACCAACAATCTAA 562
DB 61 ACATCATTAAGCAATATCAGTTACACAGCCCAACACCTCGCGAGCGCTGACCAACAATCTAA 120
QY 563 ATGAAGTCAGGACCACTTCACAGATACCTTTACCAACACACAGATGATCTGACCTCT 622
DB 121 ATGAAGTCAGGACCACTTCACAGATACCTTTACCAACACACAGATGATCTGACCTCT 180
QY 623 TGAATAATACCTCGGCCAACATCCGTTTGGATTCTGTTCTCTCAGGATGCAACAGATT 682
DB 181 TGAATAATACCTCGGCCAACATCCGTTTGGATTCTGTTCTCTCAGGATGCAACAGATT 240
QY 683 TGATGAGTCGAGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAAGAAATGA 742
DB 241 TGATGAGTCGAGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAAGAAATGA 300
QY 743 AGCTAGTAGCTCCAAAGCATGCTCAGCTCATCAAGAATTTTCAATACATACAGGTCAC 802
DB 301 AGCTAGTAGCTCCAAAGCATGCTCAGCTCATCAAGAATTTTCAATACATACAGGTCAC 360
QY 803 CCGGCCCAAGGGTCCAAAGAGTGACAGAGGATCCAGGAGCCCTCGCCCAACTGGCA 862
DB 361 CCGGCCCAAGGGTCCAAAGAGTGACAGAGGATCCAGGAGCCCTCGCCCAACTGGCA 420
QY 863 ACAAGGACAGAAAGGAGAGAGGGGAGCCCTGGACCACTCGCCCTCGGGTGAGAGAG 922
DB 421 ACAAGGACAGAAAGGAGAGAGGGGAGCCCTGGACCACTCGCCCTCGGGTGAGAGAG 480
QY 923 GCCCAATTGGACAGCTGTCCCGCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCC 982
DB 481 GCCCAATTGGACAGCTGTCCCGCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCC 540
QY 983 AGGGCCCCAAAGGCTCCCTGTTCCTTCCCTGGGAAGCCCGCCCTCAGGCCCCCAGTGGG 1042
DB 541 AGGGCCCCAAAGGCTCCCTGTTCCTTCCCTGGGAAGCCCGCCCTCAGGCCCCCAGTGGG 600
QY 1043 ACCAGGCCCCCGGGCCCAACAGGAGAGGAGTCCCGGCCCTCAGGCCCCCTCCCTG 1102
DB 601 ACCAGGCCCCCGGGCCCAACAGGAGAGGAGTCCCGGCCCTCAGGCCCCCTCCCTG 560
QY 1103 GCTTCAGGAGACTTCAGGACACCGTTGGGAGCCCTGGGTGCTCGGACCTCGGGGACTGC 1162
DB 661 GCTTCAGGAGACTTCAGGACACCGTTGGGAGCCCTGGGTGCTCGGACCTCGGGGACTGC 720
QY 1163 CAGGCTTGCTGGGTACCAAGCATGCCAGGCCCAAGGCCCGCCCGCCCTCCTGGCC 1222
DB 721 CAGGCTTGCTGGGTACCAAGCATGCCAGGCCCAAGGCCCGCCCGCCCTCCTGGCC 780
QY 1223 CATCAGGAGCGTGGTGGCCCTCGCAAGTGAAGCAACCCCGGACCGGAGGACA 1282
DB 781 CATCAGGAGCGTGGTGGCCCTCGCAAGTGAAGCAACCCCGGACCGGAGGACA 840
QY 1283 ATGGCTGCCCGCTCAGTGGAAAGACTTCACAGCAAAATGCTACTATTTTTCAGTTGAGA 1342
DB 841 ATGGCTGCCCGCTCAGTGGAAAGACTTCACAGCAAAATGCTACTATTTTTCAGTTGAGA 900
QY 1343 AGAAATTTTTCAGATGCAAGGCTTTCTGTGAGACAAAGTCTTCACATCTGTTTCA 1402
DB 901 AGAAATTTTTCAGATGCAAGGCTTTCTGTGAGACAAAGTCTTCACATCTGTTTCA 960
```

```
QY 1403 TAAACACTAGAGGAACACAGCAATGATGATAAAAAACAGATGTTAGGAGAGAGCCACT 1462
DB 961 TAAACACTAGAGGAACACAGCAATGATGATAAAAAACAGATGTTAGGAGAGAGCCACT 1020
QY 1463 GGATCGGCTCAGACTCAGAGCTCAGAGCGTGAATGATGGAAGTGGCTGGATGGGACATCTC 1522
DB 1021 GGATCGGCTCAGACTCAGAGCTCAGAGCGTGAATGATGGAAGTGGCTGGATGGGACATCTC 1080
QY 1523 CAGACTACAAAAATTTGGAAGCTGGACACCGGATAACTGGGTCATGCCATGGGCCAG 1582
DB 1081 CAGACTACAAAAATTTGGAAGCTGGACACCGGATAACTGGGTCATGCCATGGGCCAG 1140
QY 1583 GAGAAGACTGTCTGGGTTGATTTATGCTGGCAGTGGAAACGATTTCCAAATGTGAAGACG 1642
DB 1141 GAGAAGACTGTCTGGGTTGATTTATGCTGGCAGTGGAAACGATTTCCAAATGTGAAGACG 1200
QY 1643 TCAATAACTTTCATTTCCGAAAAAGACAGGAGACAGTACTGTCATCTGCATTA 1695
DB 1201 TCAATAACTTTCATTTCCGAAAAAGACAGGAGACAGTACTGTCATCTGCATTA 1253

RESULT 10
AC016128/c
LOCUS
DEFINITION
AC016128
AC016128.4 GI:10046526
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 169088)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP11-324G2
2 (bases 1 to 169088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castie,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArillano,K., Dewar,K., Domino,M., Donellan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hago,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 9, 2000 this sequence version replaced gi:6649269.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L999
Center clone name: 324_G2
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165522 bases at least Q40
```


Consensus quality: 167248 bases at least Q30
 Consensus quality: 167996 bases at least Q20
 Insert size: 177000; agarose-fp
 Insert size: 168388; sum-of-contigs
 Quality coverage: 6.2 in Q20 bases; agarose-fp
 Quality coverage: 6.5 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 17943: contig of 17943 bp in length
 * 17944 18043: gap of 100 bp
 * 18044 22424: contig of 4381 bp in length
 * 22425 22524: gap of 100 bp
 * 22525 38094: contig of 15570 bp in length
 * 38095 38194: gap of 100 bp
 * 38195 49220: contig of 11026 bp in length
 * 49221 49320: gap of 100 bp
 * 49321 66394: contig of 17074 bp in length
 * 66395 66494: gap of 100 bp
 * 66495 91692: contig of 25198 bp in length
 * 91693 91792: gap of 100 bp
 * 91793 120869: contig of 29077 bp in length
 * 120870 120969: gap of 100 bp
 * 120970 169088: contig of 48119 bp in length.

FEATURES

source
 1. .169088
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18"
 /clone="RP11-324G2"
 /clone_lib="RPC1-11 Human Male BAC"
 1. .17943
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"
 18044..22424
 /note="assembly_fragment"
 22525..38094
 /note="assembly_fragment"
 38195..49220
 /note="assembly_fragment"
 49321..66394
 /note="assembly_fragment"
 66495..91692
 /note="assembly_fragment"
 91793..120869
 /note="assembly_fragment"
 120970..169088
 /note="assembly_fragment
 clone_end:T7
 vector_side:right"
 48103 a 33537 c 34088 g 52660 t 700 others

BASE COUNT 48103 a 33537 c 34088 g 52660 t 700 others
 ORIGIN
 Query Match 47.1%; Score 798; DB 2; Length 169088;
 Best Local Similarity 100.0%; Pred. No. 3.8e-177;
 Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCACGAATCTGACGAAGTACACGCGTCTCCAGGGAATCTGCAGAACCAATGTAT 60
 Db 165666 GTCACGAATCTGACGAAGTACACGCGTCTCCAGGGAATCTGCAGAACCAATGTAT 165607
 QY 61 TCTCATATGTGTCATCATGACCTCAACACCTGAACCTGACCCAGGTGCAGCAGAGG 120
 Db 165606 TCTCATATGTGTCATCATGACCTCAACACCTGAACCTGACCCAGGTGCAGCAGAGG 165547

QY 121 AACCTCATCAGCAATCTGCAGCGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 180
 Db 165546 AACCTCATCAGCAATCTGCAGCGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 165487
 QY 181 AGAAGCAGCTTTCAAATCTGCAGCAGGTTTCTTCAAGCCAAAGAGGACAGCGATTGG 240
 Db 165486 AGAAGCAGCTTTCAAATCTGCAGCAGGTTTCTTCAAGCCAAAGAGGACAGCGATTGG 165427
 QY 241 CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCTGGCTGCCAAACACTCTGGTTGGCCAAA 300
 Db 165426 CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCTGGCTGCCAAACACTCTGGTTGGCCAAA 165367
 QY 301 GCCAACACGACACACCTGGAGGATATGAACAGCAGCTCAACTCATTCACAGTCCAGATG 360
 Db 165366 GCCAACACGACACACCTGGAGGATATGAACAGCAGCTCAACTCATTCACAGTCCAGATG 165307
 QY 361 GAGAACATCACCACCTATCTCAAGCCCAAGCAGCAGCAACCTGAAAGACCTGCGAGGACTTA 420
 Db 165306 GAGAACATCACCACCTATCTCAAGCCCAAGCAGCAGCAACCTGAAAGACCTGCGAGGACTTA 165247
 QY 421 CACAAAGATCGAGAGATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG 480
 Db 165246 CACAAAGATCGAGAGATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG 165187
 QY 481 CTCCTTTGAGACGGATATTGTGAACATCATTTAGCAATATCAGTTTACAGCCCCACCTG 540
 Db 165186 CTCCTTTGAGACGGATATTGTGAACATCATTTAGCAATATCAGTTTACAGCCCCACCTG 165127
 QY 541 CGGACCTGACCAAGCAATCTAAATGAAGTCAGACCACTTGCACAGATACCCCTTACCAA 600
 Db 165126 CGGACCTGACCAAGCAATCTAAATGAAGTCAGACCACTTGCACAGATACCCCTTACCAA 165067
 QY 601 CACACAGATGATCTGACCTCTTGAATAATACCTGGCCAAACATCGCTTTGGATTCTGTT 660
 Db 165066 CACACAGATGATCTGACCTCTTGAATAATACCTGGCCAAACATCGCTTTGGATTCTGTT 165007
 QY 661 TCTCTCAGGATCAACAAGATTTGATGAGTGCAGGTTAGACACTGAAGTAGCAACTTA 720
 Db 165006 TCTCTCAGGATCAACAAGATTTGATGAGTGCAGGTTAGACACTGAAGTAGCAACTTA 164947
 QY 721 TCAGTCAATGGAAGAAATGAAGCTAGTAGCTCAAGCATGTCAGCTCATCAAGAAT 780
 Db 164946 TCAGTCAATGGAAGAAATGAAGCTAGTAGCTCAAGCATGTCAGCTCATCAAGAAT 164887
 QY 781 TTTACAATACTACAAGGT 798
 Db 164886 TTTACAATACTACAAGGT 164869
 RESULT 11
 AP000900
 LOCUS
 DEFINITION
 Homo sapiens chromosome 18 clone RP11-683J11 map 18p11.3, WORKING
 DRAFT SEQUENCE, 20 unordered pieces.
 AP000900
 VERSION
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 Homo sapiens DNA, clone:RP11-683J11.
 SOURCE
 Homo sapiens
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 178022)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Homo sapiens 178,022 genomic DNA of 18p11.3
 Published Only in DataBase (1999)
 2 (bases 1 to 178022)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Direct Submission
 Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,

Japan (E-mail: hattori@gs.riken.go.jp,
 URL: http://hgp.gs.riken.go.jp/, Tel: 81-42-778-9923,
 Fax: 81-42-778-9924)

On May 31, 2000 this sequence version replaced gi:6997403.

COMMENT

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: http://hgp.gs.riken.go.jp/

Contact: hattori@gs.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: RP11-683J11

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 161944 bases at least Q40

Consensus quality: 170357 bases at least Q30

Consensus quality: 174322 bases at least Q20

Insert size: 176122; sum-of-contigs

Quality coverage: 4.56x in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 24283 contig of 24283 bp in length
24384 46875 contig of 22492 bp in length
46976 65058 contig of 18083 bp in length
65159 81322 contig of 16164 bp in length
81423 98823 contig of 17401 bp in length
98924 110068 contig of 11145 bp in length
110169 122716 contig of 12548 bp in length
122817 133089 contig of 10273 bp in length
133190 141206 contig of 8017 bp in length
141307 149050 contig of 7744 bp in length
149151 156105 contig of 6955 bp in length
156206 162048 contig of 5843 bp in length
162149 165118 contig of 2970 bp in length
165219 167480 contig of 2262 bp in length
167581 169986 contig of 2406 bp in length
170087 171909 contig of 1823 bp in length
172010 172317 contig of 308 bp in length
172418 174663 contig of 2246 bp in length
174764 176271 contig of 1508 bp in length
176372 178022 contig of 1651 bp in length

```

Sequence updated (24-Dec-1999)

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 24283: contig of 24283 bp in length
* 24284 24383: gap of 100 bp
* 24384 46875: contig of 22492 bp in length
* 46876 46975: gap of 100 bp
* 46976 65058: contig of 18083 bp in length
* 65059 65158: gap of 100 bp
* 65159 81322: contig of 16164 bp in length
* 81323 81422: gap of 100 bp
* 81423 98823: contig of 17401 bp in length
* 98824 98923: gap of 100 bp
* 98924 110068: contig of 11145 bp in length
* 110069 110168: gap of 100 bp
* 110169 122716: contig of 12548 bp in length
* 122717 122816: gap of 100 bp

```

```

* 122817 133089: contig of 10273 bp in length
* 133090 133189: gap of 100 bp
* 133190 141206: contig of 8017 bp in length
* 141207 141306: gap of 100 bp
* 141307 149050: contig of 7744 bp in length
* 149051 149150: gap of 100 bp
* 149151 156105: contig of 6955 bp in length
* 156106 156205: gap of 100 bp
* 156206 162048: contig of 5843 bp in length
* 162049 162148: gap of 100 bp
* 162149 165118: contig of 2970 bp in length
* 165119 165218: gap of 100 bp
* 165219 167480: contig of 2262 bp in length
* 167481 167580: gap of 100 bp
* 167581 169986: contig of 2406 bp in length
* 169987 170086: gap of 100 bp
* 170087 171909: contig of 1823 bp in length
* 171910 172009: gap of 100 bp
* 172010 172317: contig of 308 bp in length
* 172318 172417: gap of 100 bp
* 172418 174663: contig of 2246 bp in length
* 174664 174763: gap of 100 bp
* 174764 176271: contig of 1508 bp in length
* 176272 176371: gap of 100 bp
* 176372 178022: contig of 1651 bp in length.

```

FEATURES

Source

1. 178022

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="18"

/map="18p11.3"

/clone="RP11-683J11"

1. 24283

/note="assembly_fragment"

24384..46875

/note="assembly_fragment"

46976..65058

/note="assembly_fragment"

65159..81322

/note="assembly_fragment"

81423..98823

/note="assembly_fragment"

98924..110068

/note="assembly_fragment"

110169..122716

/note="assembly_fragment"

122817..133089

/note="assembly_fragment"

133190..141206

/note="assembly_fragment"

141307..149050

/note="assembly_fragment"

149151..156105

/note="assembly_fragment"

156206..162048

/note="assembly_fragment"

162149..165118

/note="assembly_fragment clone_end:SP6 vector_side:left"

165219..167480

/note="assembly_fragment"

167581..169986

/note="assembly_fragment"

170087..171909

/note="assembly_fragment"

172010..172317

/note="assembly_fragment clone_end:T7 vector_side:right"

172418..174663

/note="assembly_fragment"

174764..176271

/note="assembly_fragment"

176372..178022

/note="assembly_fragment"

BASE COUNT 53972 a 34911 c 36303 g 50935 t 1901 others

ORIGIN

Query Match 47.1%; Score 798; DB 2; Length 178022;
Best Local Similarity 100.0%; Pred. No. 3.8e-177;
Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCACGAATCTGCAGCAAGATACACGGTGCTCCAGGCAATCTGCAGAACCAATGTAT 60
|||||
Db 13104 GTCACGAATCTGCAGCAAGATACACGGTGCTCCAGGCAATCTGCAGAACCAATGTAT 13163
|||||
QY 61 TCTCATTAATCTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAGCAGAGG 120
|||||
Db 13164 TCTCATTAATCTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAGCAGAGG 13223
|||||
QY 121 AACCTCATACGAATCTGCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 180
|||||
Db 13224 AACCTCATACGAATCTGCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 13283
|||||
QY 181 AAGAACGACTTTCAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGAGGACCGATTGG 240
|||||
Db 13284 AAGAACGACTTTCAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGAGGACCGATTGG 13343
|||||
QY 241 CTGAAGGAGAAAGTGCAGAGCTTGCAGACGCTGGCTGCCAACCAACTCTGGTTGGCCAAA 300
|||||
Db 13344 CTGAAGGAGAAAGTGCAGAGCTTGCAGACGCTGGCTGCCAACCAACTCTGGTTGGCCAAA 13403
|||||
QY 301 GCCAACAGACACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGTCAGATG 360
|||||
Db 13404 GCCAACAGACACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGTCAGATG 13463
|||||
QY 361 GAGAACTACCACTATCTCTCAAGCCAGGACGACGAACCTGAAAGACCTGCAGGACTTA 420
|||||
Db 13464 GAGAACTACCACTATCTCTCAAGCCAGGACGACGAACCTGAAAGACCTGCAGGACTTA 13523
|||||
QY 421 CACAAAGATCGAGAGATAGAAGCGCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG 480
|||||
Db 13524 CACAAAGATCGAGAGATAGAAGCGCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG 13583
|||||
QY 481 CTCTTTGAGACGGATATTGTGAACATCATTTAGCAATATCAGTTTACAGCCCCACCCCTG 540
|||||
Db 13584 CTCTTTGAGACGGATATTGTGAACATCATTTAGCAATATCAGTTTACAGCCCCACCCCTG 13643
|||||
QY 541 CGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGCACATACCCCTTACCATA 600
|||||
Db 13644 CGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGCACATACCCCTTACCATA 13703
|||||
QY 601 CACACAGATGATCTGACCTCTCTTGAATAATACCTTGCCCAACATCCGTTTGGATTCTGTT 660
|||||
Db 13704 CACACAGATGATCTGACCTCTCTTGAATAATACCTTGCCCAACATCCGTTTGGATTCTGTT 13763
|||||
QY 661 TCTCTCAGGATGCAACAAGATTTGATGAGGTGAGGTTAGACACTGAAGTAGCCAACTTA 720
|||||
Db 13764 TCTCTCAGGATGCAACAAGATTTGATGAGGTGAGGTTAGACACTGAAGTAGCCAACTTA 13823
|||||
QY 721 TCAGTGATTATGAAGAAATGAAGCTAGTAGACTTCAAGCATGGTCAGCTCATCAAGAAT 780
|||||
Db 13824 TCAGTGATTATGAAGAAATGAAGCTAGTAGACTTCAAGCATGGTCAGCTCATCAAGAAT 13883
|||||
QY 781 TTTTACAACTACTACAAGT 798
|||||
Db 13884 TTTTACAACTACTACAAGT 13901
|||||

RESULT 12
AP001022
LOCUS AP001022 187635 bp DNA linear HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-815L4 map 18p11.3, WORKING
DRAFT SEQUENCE, 33 unordered pieces.
AP001022
ACCESSION AP001022.2 GI:8117692
VERSION HTG; HTGS; PHASE1; HTGS; DRAFT.
KEYWORDS Homo sapiens DNA, clone:RP11-815L4.
SOURCE
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 187635)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 187,635 genomic DNA of 18p11.3
Published Only in DataBase (2000)
2 (bases 1 to 187635)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (05-JAN-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan [E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6997772.
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: Humdraft18
Center clone name: RP11-815L4
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167192 bases at least Q40
Consensus quality: 177020 bases at least Q30
Consensus quality: 182105 bases at least Q20
Insert size: 184435; sum-of-contigs
Quality coverage: 4.35x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
33 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1 21332 contig of 21332 bp in length
21433 42072 contig of 20640 bp in length
42173 54336 contig of 12164 bp in length
54437 66766 contig of 12330 bp in length
66867 77849 contig of 10983 bp in length
77950 87798 contig of 9849 bp in length
87899 99510 contig of 11612 bp in length
99611 107273 contig of 7663 bp in length
107374 113189 contig of 5816 bp in length
113290 120559 contig of 7270 bp in length
120660 126468 contig of 5809 bp in length
126569 13619 contig of 5051 bp in length
131720 136810 contig of 5091 bp in length
136911 139925 contig of 3015 bp in length
140026 144232 contig of 4207 bp in length
144333 148902 contig of 4570 bp in length
149003 152918 contig of 3916 bp in length
153019 156399 contig of 3381 bp in length
156500 158933 contig of 2434 bp in length
159034 161884 contig of 2851 bp in length
161985 164301 contig of 2317 bp in length
164402 166300 contig of 1899 bp in length
166401 168580 contig of 2180 bp in length
168681 171270 contig of 2590 bp in length
171371 174092 contig of 2722 bp in length
174193 176095 contig of 1903 bp in length
176196 177890 contig of 1695 bp in length
177991 179786 contig of 1796 bp in length
179887 181478 contig of 1592 bp in length
181579 182895 contig of 1317 bp in length

182996 184841 contig of 1846 bp in length
184942 186345 contig of 1404 bp in length
186446 187635 contig of 1190 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 21332: contig of 21332 bp in length
* 21333 21432: gap of 100 bp
* 21433 42072: contig of 20640 bp in length
* 42073 42172: gap of 100 bp
* 42173 54336: contig of 12164 bp in length
* 54337 54436: gap of 100 bp
* 54437 66766: contig of 12330 bp in length
* 66767 66866: gap of 100 bp
* 66867 77849: contig of 10983 bp in length
* 77850 77949: gap of 100 bp
* 77950 87798: contig of 9849 bp in length
* 87799 87898: gap of 100 bp
* 87899 99510: contig of 11612 bp in length
* 99511 99610: gap of 100 bp
* 99611 107273: contig of 7663 bp in length
* 107274 107373: gap of 100 bp
* 107374 113189: contig of 5816 bp in length
* 113190 113289: gap of 100 bp
* 113290 120559: contig of 7270 bp in length
* 120560 120659: gap of 100 bp
* 120660 126468: contig of 5809 bp in length
* 126469 126568: gap of 100 bp
* 126569 131619: contig of 5051 bp in length
* 131620 131719: gap of 100 bp
* 131720 136810: contig of 5091 bp in length
* 136811 136910: gap of 100 bp
* 136911 139925: contig of 3015 bp in length
* 139926 140025: gap of 100 bp
* 140026 144232: contig of 4207 bp in length
* 144233 144332: gap of 100 bp
* 144333 148902: contig of 4570 bp in length
* 148903 149002: gap of 100 bp
* 149003 152918: contig of 3916 bp in length
* 152919 153018: gap of 100 bp
* 153019 156399: contig of 3381 bp in length
* 156400 156499: gap of 100 bp
* 156500 158933: contig of 2434 bp in length
* 158934 159033: gap of 100 bp
* 159034 161884: contig of 2851 bp in length
* 161885 161984: gap of 100 bp
* 161985 164301: contig of 2317 bp in length
* 164302 164401: gap of 100 bp
* 164402 166300: contig of 1899 bp in length
* 166301 166400: gap of 100 bp
* 166401 168580: contig of 2180 bp in length
* 168581 168680: gap of 100 bp
* 168681 171270: contig of 2590 bp in length
* 171271 171370: gap of 100 bp
* 171371 174092: contig of 2722 bp in length
* 174093 174192: gap of 100 bp
* 174193 176095: contig of 1903 bp in length
* 176096 176195: gap of 100 bp
* 176196 177890: contig of 1695 bp in length
* 177891 177990: gap of 100 bp
* 177991 179786: contig of 1796 bp in length
* 179787 179886: gap of 100 bp
* 179887 181478: contig of 1592 bp in length
* 181479 181578: gap of 100 bp
* 181579 182895: contig of 1317 bp in length
* 182896 182995: gap of 100 bp
* 182996 184841: contig of 1846 bp in length

* 184842 184941: gap of 100 bp
* 184942 186345: contig of 1404 bp in length
* 186346 186445: gap of 100 bp
* 186446 187635: contig of 1190 bp in length.

FEATURES

source

Location/Qualifiers
1..187635
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-815L4"
1..21332
/note="assembly_fragment"
21433..42072
/note="assembly_fragment"
42173..54336
/note="assembly_fragment"
54437..66766
/note="assembly_fragment"
66867..77849
/note="assembly_fragment"
77950..87798
/note="assembly_fragment clone_end:T7 vector_side:right"
87899..99510
/note="assembly_fragment"
99611..107273
/note="assembly_fragment"
107374..113189
/note="assembly_fragment clone_end:SP6 vector_side:left"
113290..120559
/note="assembly_fragment"
120660..126468
/note="assembly_fragment"
126569..131619
/note="assembly_fragment"
131720..136810
/note="assembly_fragment"
136911..139925
/note="assembly_fragment"
140026..144232
/note="assembly_fragment"
144333..148902
/note="assembly_fragment"
149003..152918
/note="assembly_fragment"
153019..156399

Query Match 47.1%; Score 798; DB 2; Length 187635;
Best Local Similarity 100.0%; Pred. No. 3.8e-177;
Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCACGAATCTGCAGCAAGATACCAGCGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
|||||
Db 58354 GTCACGAATCTGCAGCAAGATACCAGCGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 58413
QY 61 TCTCATAATGTGGTTCATCATGAACCTCAACAACTGAACCTGACCCAGGTGCAGCAGAG 120
|||||
Db 58414 TCTCATAATGTGGTTCATCATGAACCTCAACAACTGACCCAGGTGCAGCAGAG 58473
QY 121 AACCTCATCAGNAATCTGCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 180
|||||
Db 58474 AACCTCATCAGNAATCTGCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 58533
QY 181 AGAAGCAGCTTTCAAATCTGCAGCAGGTTTTTTTCAAGCCCAAGAGGACGCGATTGG 240
|||||
Db 58534 AGAAGCAGCTTTCAAATCTGCAGCAGGTTTTTTTCAAGCCCAAGAGGACGCGATTGG 58593
QY 241 CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCGTGGCTGCCAACAACTCTGCGTTGGCCAA 300
|||||
Db 58594 CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCGTGGCTGCCAACAACTCTGCGTTGGCCAA 58653
QY 301 GCCAACACGACACCTCGGAGGATATGACAGCCAGCTCAACTCACTTCACAGTTCAGATG 360
|||||

| | | | |
|-------------|---|---|----------------------------|
| Db | 58654 | GCCAAACACACACCCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGTCAGATG | 58713 |
| QY | 361 | GAGAACATCACCCTATCTCTCAGCCACGAGCAGAAACCTGAAAGACCTGCAGGACTTA | 420 |
| Db | 58714 | GAGAACATCACCCTATCTCTCAGCCACGAGCAGAAACCTGAAAGACCTGCAGGACTTA | 58773 |
| QY | 421 | CACAAAGATCGAGAGATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG | 480 |
| Db | 58774 | CACAAAGATCGAGAGATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG | 58833 |
| QY | 481 | CTCTTTGAGACGGATATGTGAACATCATTAGCAATATCAGTTACACAGCCACCACTG | 540 |
| Db | 58834 | CTCTTTGAGACGGATATGTGAACATCATTAGCAATATCAGTTACACAGCCACCACTG | 58893 |
| QY | 541 | CGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAA | 600 |
| Db | 58894 | CGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAA | 58953 |
| QY | 601 | CACACAGATGATCTGACCTCCTTGAATATACCTCGCCAAACATCCGTTTGGATTCTGTT | 660 |
| Db | 58954 | CACACAGATGATCTGACCTCCTTGAATATACCTCGCCAAACATCCGTTTGGATTCTGTT | 59013 |
| QY | 661 | TCCTCAGGATGCAACAAGATTTGATCAGTCGAGCTTAGACACTGAAGTAGCCAATTA | 720 |
| Db | 59014 | TCCTCAGGATGCAACAAGATTTGATCAGTCGAGCTTAGACACTGAAGTAGCCAATTA | 59073 |
| QY | 721 | TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTTCAAGCATGGTCAGCTCATCAAGAAT | 780 |
| Db | 59074 | TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTTCAAGCATGGTCAGCTCATCAAGAAT | 59133 |
| QY | 781 | TTTACAATACTACAAGGT | 798 |
| Db | 59134 | TTTACAATACTACAAGGT | 59151 |
| RESULT 13 | | | |
| AP000915/c | | | |
| LOCUS | AP000915 | 188439 bp | DNA linear PRI 27-APR-2002 |
| DEFINITION | Homo sapiens genomic DNA, chromosome 18p clone:RP11-720L2, complete sequences. | | |
| ACCESSION | AP000915 | | |
| VERSION | AP000915.5 | GI:20334314 | |
| KEYWORDS | HTG. | | |
| SOURCE | Homo sapiens DNA, clone:RP11-720L2. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| AUTHORS | Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y. | | |
| TITLE | Homo sapiens genomic DNA | | |
| JOURNAL | Published Only in Database (1999) | | |
| REFERENCE | 2 (bases 1 to 188439) | | |
| AUTHORS | Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical and Chemical research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170) | | |
| COMMENT | On Apr 26, 2002 this sequence version replaced gi:9188470. | | |
| FEATURES | Location/Qualifiers | | |
| source | 1. .188439 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /chromosome="18" | | |
| | /map="18p" | | |
| | /clone="RP11-720L2" | | |
| BASE COUNT | 54003 a | 41134 c | 40027 g 53275 t |
| ORIGIN | | | |
| Query Match | 47.1%; | Score 798; | DB 9; Length 188439; |

| | | | |
|--|--|--|----------------------------|
| Best Local Similarity 100.0%; Pred. No. 3.8e-177; | | | |
| Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | | |
| QY | 1 | GTCCACGAATCTGCAGCAAGATACCGCTGCTCCAGGGCAATCTGCAGAACCAAAATGTAT | 60 |
| Db | 56796 | GTCCAGNATCTGCAGCAAGATACCGCTGCTCCAGGGCAATCTGCAGAACCAAAATGTAT | 56737 |
| QY | 61 | TCTCATAATGTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTCAGCAGAGG | 120 |
| Db | 56736 | TCTCATAATGTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTCAGCAGAGG | 56677 |
| QY | 121 | AACCTCATCAGGAATCTGCAGCGTCTGTGGATGACACAAGCCAGGCTATCCAGCGAATC | 180 |
| Db | 56676 | AACCTCATCAGGAATCTGCAGCGTCTGTGGATGACACAAGCCAGGCTATCCAGCGAATC | 56617 |
| QY | 181 | AAGAAGCACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAGAGGACACGGATTGG | 240 |
| Db | 56616 | AAGAAGCACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAGAGGACACGGATTGG | 56557 |
| QY | 241 | CTGAAGGAGAAAGTCGAGAGCTTTGCAGAGCGCTGGCTGCCAACAACTCTGCGTTGGCCAAA | 300 |
| Db | 56556 | CTGAAGGAGAAAGTCGAGAGCTTTGCAGAGCGCTGGCTGCCAACAACTCTGCGTTGGCCAAA | 56497 |
| QY | 301 | GCCAAACAGACACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGTCAGATG | 360 |
| Db | 56496 | GCCAAACAGACACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGTCAGATG | 56437 |
| QY | 361 | GAGAAATCACCACCTATCTCTCAAGCCACAGCAGCAACCTGAAAGACCTGCGAGACTTA | 420 |
| Db | 56436 | GAGAAATCACCACCTATCTCTCAAGCCACAGCAGCAACCTGAAAGACCTGCGAGACTTA | 56377 |
| QY | 421 | CACAAAGATCGAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG | 480 |
| Db | 56376 | CACAAAGATCGAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG | 56317 |
| QY | 481 | CTCTTTGAGACGGATATGTGAACATCATTAGCAATATCAGTTACACAGCCCACTGCTG | 540 |
| Db | 56316 | CTCTTTGAGACGGATATGTGAACATCATTAGCAATATCAGTTACACAGCCCACTGCTG | 56257 |
| QY | 541 | CGGACCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAA | 600 |
| Db | 56256 | CGGACCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCCCTTACCAA | 56197 |
| QY | 601 | CACACAGATGATCGACCTCTTGAATAATACCTGGCCAACTCCGTTTGGATTCTGTT | 660 |
| Db | 56196 | CACACAGATGATCGACCTCTTGAATAATACCTGGCCAACTCCGTTTGGATTCTGTT | 56137 |
| QY | 661 | TCTCTCAGGATGCAACAAGATTTGATGAGGTCGAGGCTAGACACTGAAGTAGCCAATTA | 720 |
| Db | 56136 | TCTCTCAGGATGCAACAAGATTTGATGAGGTCGAGGCTAGACACTGAAGTAGCCAATTA | 56077 |
| QY | 721 | TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTTCAAGCATGGTCAGCTCATCAAGAAT | 780 |
| Db | 56076 | TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTTCAAGCATGGTCAGCTCATCAAGAAT | 56017 |
| QY | 781 | TTTACAATACTACAAGGT | 798 |
| Db | 56016 | TTTACAATACTACAAGGT | 55999 |
| . | | | |
| RESULT | 14 | | |
| LOCUS | AP000939 | 188255 bp | DNA linear HTG 30-MAY-2000 |
| DEFINITION | Homo sapiens chromosome 18 clone RP11-839023 map 18p11.3, WORKING DRAFT SEQUENCE, 32 unordered pieces. | | |
| ACCESSION | AP000939 | | |
| VERSION | AP000939.3 GI:8119080 | | |
| KEYWORDS | HTG; HTGS_PHASE1; HTGS_DRAFT. | | |
| SOURCE | Homo sapiens DNA, clone:RP11-839023. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| | 1 (bases 1 to 188255) | | |

AUTHORS

Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Homo sapiens 188,255 genomic DNA of 18p11.3
 Published Only in Database (1999)
 2 (bases 1 to 188255)
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
 Japan (E-mail: hattori@gsc.riken.go.jp,
 URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,
 Fax: 81-42-778-9924)
 On May 31, 2000 this sequence version replaced gi:6997751.

COMMENT

----- Genomic Center
 Center: RIKEN Genomic Sciences Center (GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft18
 Center clone name: RP11-839023
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 171068 bases at least Q40
 Consensus quality: 178865 bases at least Q30
 Consensus quality: 182445 bases at least Q20
 Insert size: 185155; sum-of-contigs
 Quality coverage: 4.41x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 16096 contig of 16096 bp in length
 16197 33789 contig of 17593 bp in length
 33890 51010 contig of 17121 bp in length
 51111 63558 contig of 12448 bp in length
 63659 75056 contig of 11398 bp in length
 75157 84737 contig of 9581 bp in length
 84838 90909 contig of 6072 bp in length
 91010 98615 contig of 7606 bp in length
 98716 105978 contig of 6971 bp in length
 106079 113049 contig of 5388 bp in length
 113150 119201 contig of 4968 bp in length
 119302 126582 contig of 5817 bp in length
 126683 132070 contig of 5388 bp in length
 132171 137138 contig of 4968 bp in length
 137239 143055 contig of 5817 bp in length
 143156 148639 contig of 5484 bp in length
 148740 153086 contig of 4347 bp in length
 153087 153186 contig of 100 bp in length
 153187 158921 contig of 5735 bp in length
 158922 159021 contig of 100 bp in length
 159022 163174 contig of 4153 bp in length
 163175 163274 contig of 100 bp in length
 163275 167066 contig of 3792 bp in length
 167067 167166 contig of 100 bp in length
 167167 170120 contig of 2954 bp in length
 170121 170220 contig of 100 bp in length
 170221 172430 contig of 2210 bp in length
 172431 172530 contig of 100 bp in length
 172531 175598 contig of 3068 bp in length
 175599 175698 contig of 100 bp in length
 175699 177118 contig of 1420 bp in length
 177119 177218 contig of 100 bp in length
 177219 178531 contig of 1313 bp in length
 178532 178631 contig of 100 bp in length
 178632 179857 contig of 1226 bp in length
 179858 179957 contig of 100 bp in length
 179958 181064 contig of 1107 bp in length
 181065 181164 contig of 100 bp in length
 181165 182735 contig of 1571 bp in length
 182736 182835 contig of 100 bp in length
 182836 183923 contig of 1088 bp in length
 183924 184023 contig of 100 bp in length
 184024 185810 contig of 1787 bp in length
 185811 185910 contig of 100 bp in length
 185911 187145 contig of 1235 bp in length
 187146 187245 contig of 100 bp in length
 187246 188255 contig of 1010 bp in length

Sequence updated (06-Jan-2000)

FEATURES

Location/Qualifiers

Sequence updated (26-May-2000).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 16096: contig of 16096 bp in length
 * 16097 16196: gap of 100 bp
 * 16197 33789: contig of 17593 bp in length
 * 33790 33889: gap of 100 bp
 * 33890 51010: contig of 17121 bp in length
 * 51011 51110: gap of 100 bp
 * 51111 63558: contig of 12448 bp in length
 * 63559 63658: gap of 100 bp
 * 63659 75056: contig of 11398 bp in length
 * 75057 75156: gap of 100 bp
 * 75157 84737: contig of 9581 bp in length
 * 84738 84837: gap of 100 bp
 * 84838 90909: contig of 6072 bp in length
 * 90910 91009: gap of 100 bp
 * 91010 98615: contig of 7606 bp in length
 * 98616 98715: gap of 100 bp
 * 98716 105978: contig of 7263 bp in length
 * 105979 106078: gap of 100 bp
 * 106079 113049: contig of 6971 bp in length
 * 113050 113149: gap of 100 bp
 * 113150 119201: contig of 6052 bp in length
 * 119202 119301: gap of 100 bp
 * 119302 126582: contig of 7281 bp in length
 * 126583 126682: gap of 100 bp
 * 126683 132070: contig of 5388 bp in length
 * 132071 132170: gap of 100 bp
 * 132171 137138: contig of 4968 bp in length
 * 137139 137238: gap of 100 bp
 * 137239 143055: contig of 5817 bp in length
 * 143056 143155: gap of 100 bp
 * 143156 148639: contig of 5484 bp in length
 * 148640 148739: gap of 100 bp
 * 148740 153086: contig of 4347 bp in length
 * 153087 153186: gap of 100 bp
 * 153187 158921: contig of 5735 bp in length
 * 158922 159021: gap of 100 bp
 * 159022 163174: contig of 4153 bp in length
 * 163175 163274: gap of 100 bp
 * 163275 167066: contig of 3792 bp in length
 * 167067 167166: gap of 100 bp
 * 167167 170120: contig of 2954 bp in length
 * 170121 170220: gap of 100 bp
 * 170221 172430: contig of 2210 bp in length
 * 172431 172530: gap of 100 bp
 * 172531 175598: contig of 3068 bp in length
 * 175599 175698: gap of 100 bp
 * 175699 177118: contig of 1420 bp in length
 * 177119 177218: gap of 100 bp
 * 177219 178531: contig of 1313 bp in length
 * 178532 178631: gap of 100 bp
 * 178632 179857: contig of 1226 bp in length
 * 179858 179957: gap of 100 bp
 * 179958 181064: contig of 1107 bp in length
 * 181065 181164: gap of 100 bp
 * 181165 182735: contig of 1571 bp in length
 * 182736 182835: gap of 100 bp
 * 182836 183923: contig of 1088 bp in length
 * 183924 184023: gap of 100 bp
 * 184024 185810: contig of 1787 bp in length
 * 185811 185910: gap of 100 bp
 * 185911 187145: contig of 1235 bp in length
 * 187146 187245: gap of 100 bp
 * 187246 188255: contig of 1010 bp in length.

REFERENCE AUTHORS

Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 182029)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campoliano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Deaerallano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Katat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehocsky,J., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,F., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Trigliio,D.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (10-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 10, 2002 this sequence version replaced gi:21328559.
All repeats were identified using RepeatMasker:
Smt, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

| TITLE | JOURNAL | COMMENT |
|-------|---------|---------|
|-------|---------|---------|

Center: Whichead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24628
Center clone name: 213_K_19

----- Summary Statistics

Sequencing vector: pUC19; Plasmid: n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179506 bases at least Q40
Consensus quality: 180723 bases at least Q30
Consensus quality: 181030 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 181199; sum-of-contigs
Quality coverage: 8.0 in Q20 bases; agarose-fp
Quality coverage: 8.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.
 * This sequence will be replaced

- * this sequence will be replaced
- * by the finished sequence as soon as it is available and
- * the accession number will be preserved.

| | |
|-------|--------------------------------------|
| 1 | 991: contig of 951 bp in length |
| 952 | 1051: gap of 100 bp |
| 1052 | 3806: contig of 2755 bp in length |
| 3807 | 3906: gap of 100 bp |
| 3907 | 8709: contig of 4803 bp in length |
| 8710 | 8809: gap of 100 bp |
| 8810 | 17263: contig of 8454 bp in length |
| 17264 | 17363: gap of 100 bp |
| 17364 | 36938: contig of 19575 bp in length |
| 36939 | 37038: gap of 100 bp |
| 37039 | 53226: contig of 16188 bp in length |
| 53227 | 53262: gap of 100 bp |
| 53262 | 695752: contig of 16246 bp in length |

FEATURES

* 69573 69672: gap of 100 bp
 * 69673 89021: contig of 19349 bp in length
 * 89022 89121: gap of 100 bp
 * 89122 128310: contig of 39189 bp in length
 * 128311 128410: gap of 100 bp
 * 128411 182029: contig of 53619 bp in length.

| Location/Qualifiers | source |
|---------------------|--------|
| 1. .182029 | |

```

source
1. 100029
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /clone="RP24-213K19"
  /clone_lib="RPCI-24 Male Mouse BAC"
1. 951
misc feature

```

misc_feature

```

"msc_recurse
1: 1001
/note="assembly_fragment"

```

misc_feature

```

"note": "assembly_fragment"

```

misc_feature

```

/note="assembly_fragment"

```

`misc_feature`

```
/note="assembly_fragment"
```

misc_feature

```
/note="assembly_fragment"
```

misc_feature

```
/note="assembly_fragment"
```

misc_feature

```

- file=50073
- note="assembly_fragment"
- file=50073
- note="assembly_fragment"

```

misc_feature

```

info feature
/note="assembly_fragment
00122 120210

```

misc_feature

```

misc feature 128411 192029
/note="assembly_fragment

```

musc_reale

| BASE COUNT | 54245 a | 37484 c | 36555 g | 52844 t | 901 others |
|---------------------------|---------|---------|---------|---------|------------|
| ORIGIN | | | | | |
| /note="assembly_fragment" | | | | | |

```
Query Match      39.1%; Score 662; DB 2; Length 182029;
Best Local Similarity 89.3%; Pred. No. 4.7e-145;
Matches 713; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
```

| | | | |
|----|------|--|------|
| Qy | 1 | GTCACGAATCTGCAGCAAGATATACACAGCGTGTCTCAGGGCAATCTGCAGAACCAAAATGTAT | 60 |
| Db | 1931 | GTCAAAATCTGCACAAGATACTAGTGTGTCTCCAGGGCAATCTGCAGAGCAAAATGTAT | 1872 |
| Qy | 61 | TCTCATATGTGGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTCGACGAGG | 120 |
| Db | 1871 | TCTCAGAGCGTGGTTATCATGAACCTCAACAACCTGAACCTTAACCCAGGTTTCAGCAGAGG | 1812 |
| Qy | 121 | AACCTCATCATCGAATCTGCAGCGGTCTGTGGATGACACAAAGCCAGGCTATCCAGCGCAATC | 180 |
| Db | 1811 | AACCTTTATCTCAAAATCTGCAGCAGCTCTGTGGATGACACAAAGCCGTGGCCATCCAGCGCAATT | 1752 |
| Qy | 181 | AAGAACACACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAGAGGACACGGATTGG | 240 |
| Db | 1751 | AAGAAATGATTTTCAAAAATCTGCAACAGGTTTTTCTTCAAGCCAAGAGGACACCGATTGG | 1692 |
| Qy | 241 | CTGAAGCAGAGAAGTGCAGAGCTTGCAGAGCGTGCCTGCAACAACCTGCTGCTTGGCCAAA | 300 |
| Db | 1691 | CTAAGGAGAAAGTACAGAGCTTGCAGACATTGCTTGCACAACACTTGCCTCTGGCCAAA | 1632 |
| Qy | 301 | GCCAAACACGACACCCCTGGAGGATATGAACAGCCAGCTCAACTCATTTACAGGTCAGATG | 360 |
| Db | 1631 | GCCAACAATGACACCCCTAGAGGATATGAATAGCCAGCTCAGCTCATTTACAGGTCAGATG | 1572 |
| Qy | 361 | GAGAACATCACCATACTCTCAAGCCACAGCAGAGCAACCTGAAGAGCCTGCAGGACTTA | 420 |
| Db | 1571 | GACAACATTTACCACTATCTCAGAGGCCAACGAGCAGAGCCTTGAAGAGCCTTCAGGACTTA | 1512 |
| Qy | 421 | CACAAGATGACAGAGAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACGCTTCCAG | 480 |
| Db | 1511 | CACAAGATACAGAAAATAGAACAGCTGTCAAGTTCAAGCCAACTTGAAGGAACGCTTCCAG | 1452 |
| Qy | 481 | CTCTTTTGAGCGGATATGTTGAACATCATTTAGCAATATCAGTTTACAGGCCACCACTGT | 540 |

```

Db 1451 GTCTTTGAGACAGATATTGTGAACATCATAGCAACATCAGCTACACAGCCCATCACCTG 1392
QY 541 CGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGACACAGATACCCCTTACCAAA 600
Db 1391 AGGACACTGACCAAGCAATCTGAATGATGTTAGGACCAATGCACACACCTTGACCAGA 1332
QY 601 CACACAGATGATCTGACCTCCTTTGAATAATACCCCTGGCCAAACATCCGTTTGGATTCTGTT 660
Db 1331 CACACGGATGACCTGACCTCCTTTGAATAACACACTAGTCAACATCCGCTTGGATTCTATT 1272
QY 661 TCTCTCAGGATGCAACAAGATTGTGATGAGTGCAGGTTAGACACTGAAGTAGCCAACTTA 720
Db 1271 TCTCTCAGGATGCAACAAGATTGTGATGAGTGCAGGTTAGACACTGAAGTAGCCAACTTA 1212
QY 721 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAT 780
Db 1211 TCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAT 1152
QY 781 TTTACAATACTACAAGGT 798
Db 1151 TTTACCATTCTACAAGGT 1134

```

Search completed: March 21, 2003, 06:24:31
Job time : 5262.67 secs

GenCore version 5.1.4.p5_4578
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:57:56 ; Search time 258.28 Seconds
 (without alignments)
 5089.177 Million cell updates/sec

Title: US-09-763-712A-1_COPY_1_1695

Perfect score: 1695
 Sequence: 1 gtcacgaatctgcagcaaga.....cagtactgtcatctgcatta 1695

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 38737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Length | DB ID | Description |
|------------|-------------|--------|---------|--------------------|
| 1 | 1681.4 | 99.2 | 2930 10 | US-09-745-763-198 |
| 2 | 111 | 6.5 | 6728 10 | US-09-954-456-782 |
| 3 | 111 | 6.5 | 6728 10 | US-09-880-107-3946 |
| 4 | 110.6 | 6.5 | 1133 9 | US-09-924-340-57 |
| 5 | 110.6 | 6.5 | 1133 9 | US-09-924-340-57 |
| 6 | 109.4 | 6.5 | 2691 10 | US-09-925-302-64 |
| 7 | 107.2 | 6.3 | 2192 10 | US-09-925-301-42 |
| 8 | 107 | 6.3 | 5086 10 | US-09-880-107-3947 |
| 9 | 107 | 6.3 | 5145 9 | US-09-925-299-206 |
| 10 | 107 | 6.3 | 5145 10 | US-09-925-299-206 |
| 11 | 107 | 6.3 | 5432 12 | US-10-044-090-22 |
| 12 | 106.8 | 6.3 | 4908 9 | US-10-001-887-33 |
| 13 | 106 | 6.3 | 2542 9 | US-09-954-531-961 |
| 14 | 106 | 6.3 | 2542 10 | US-09-954-531-961 |
| 15 | 105.4 | 6.2 | 5416 10 | US-09-954-456-786 |
| 16 | 105.4 | 6.2 | 5416 10 | US-09-880-107-2094 |
| 17 | 105.2 | 6.2 | 6158 10 | US-09-919-497-6 |
| 18 | 105.2 | 6.2 | 6158 10 | US-09-954-456-762 |
| 19 | 104.2 | 6.1 | 4149 12 | US-10-044-090-104 |

| | | | | |
|----|-------|-----|---------|--------------------|
| 20 | 102.4 | 6.0 | 1486 10 | US-09-925-302-247 |
| 21 | 101.6 | 6.0 | 3226 10 | US-09-954-456-725 |
| 22 | 101.4 | 6.0 | 3690 12 | US-10-044-090-448 |
| 23 | 98.6 | 5.8 | 2158 9 | US-10-001-835-98 |
| 24 | 97.8 | 5.8 | 2520 10 | US-09-880-107-3685 |
| 25 | 95.8 | 5.7 | 1485 10 | US-09-925-302-246 |
| 26 | 95.2 | 5.6 | 1797 9 | US-09-978-295A-613 |
| 27 | 95.2 | 5.6 | 1797 9 | US-09-978-697-613 |
| 28 | 95.2 | 5.6 | 1797 9 | US-09-978-192A-613 |
| 29 | 95.2 | 5.6 | 1797 9 | US-09-999-832A-613 |
| 30 | 95.2 | 5.6 | 1797 9 | US-09-978-189-613 |
| 31 | 95.2 | 5.6 | 1797 9 | US-10-174-590-331 |
| 32 | 95.2 | 5.6 | 1797 9 | US-10-176-758-331 |
| 33 | 95.2 | 5.6 | 1797 9 | US-10-175-737-331 |
| 34 | 95.2 | 5.6 | 1797 9 | US-10-173-706-331 |
| 35 | 95.2 | 5.6 | 1797 9 | US-10-175-738-331 |
| 36 | 95.2 | 5.6 | 1797 9 | US-10-175-752-331 |
| 37 | 95.2 | 5.6 | 1797 9 | US-10-176-482-331 |
| 38 | 95.2 | 5.6 | 1797 9 | US-10-176-757-331 |
| 39 | 95.2 | 5.6 | 1797 9 | US-10-176-913-331 |
| 40 | 95.2 | 5.6 | 1797 9 | US-10-180-552-331 |
| 41 | 95.2 | 5.6 | 1797 9 | US-10-180-557-331 |
| 42 | 95.2 | 5.6 | 1797 9 | US-10-173-700-331 |
| 43 | 95.2 | 5.6 | 1797 9 | US-10-174-572-331 |
| 44 | 95.2 | 5.6 | 1797 9 | US-10-174-579-331 |
| 45 | 95.2 | 5.6 | 1797 9 | US-10-174-582-331 |

ALIGNMENTS

RESULT 1

US-09-745-763-198
 ; Sequence 198 Application US/09745763
 ; Patent No. US20020065394A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth
 McCoy, John M.
 Lavallie, Edward R.
 Collins-Racie, Lisa A.
 Evans, Cheryl
 Merberg, David
 Treacy, Maurice
 Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION NUMBER:

APPLICATION NUMBER: US/09/745,763

FILING DATE: 18-Jun-2000

CLASSIFICATION: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.

REGISTRATION NUMBER: 41,323

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8284

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 198:

SEQUENCE CHARACTERISTICS:

LENGTH: 2930 base pairs

TYPE: nucleic acid

```
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-09-745-763-198

Query Match      99.2%; Score 1681.4; DB 10; Length 2930;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1693; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1  GTCCACCAATCTGCAGCAAGTACCACGGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
Db 598 GTCCACCAATCTGCAGCAAGTACCACGGTCTCCAGGGCAATCTGCAGAACCAAAATGTAT 60
QY 61  TCTCATAATGTGTCTATCATGAACCTCAACAACCTCAACCTGACCCAGGTGCAGCAGAGG 120
Db 658 TCTCATAATGTGTCTATCATGAACCTCAACAACCTCAACCTGACCCAGGTGCAGCAGAGG 120
QY 121 AACCTCATCAGCAATCTGCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 180
Db 717 AACCTCATCAGCAATCTGCAGCGGTCTGTGGATGACACAGCCAGGCTATCCAGCGAATC 180
QY 181 AAGAACGACTTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGGACACGGATTGG 240
Db 777 AAGAACGACTTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGGACACGGATTGG 240
QY 241 CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCTGGCTGCCAACAACTCTCGTTGGCCAAA 300
Db 837 CTGAAGGAGAAAGTGCAGAGCTTGCAGAGCTGGCTGCCAACAACTCTCGTTGGCCAAA 300
QY 301 GCCAACAGACACCTTGAGGATATGACAGCCAGCTCAACTCATTCAGGTCTAGATG 360
Db 897 GCCAACAGACACCTTGAGGATATGACAGCCAGCTCAACTCATTCAGGTCTAGATG 360
QY 361 GAGAACATCACCACCTATCTCTCAAGCCCAAGCAGCAGCAACCTGAAAGACCTGCAGGACTTA 420
Db 957 GAGAACATCACCACCTATCTCTCAAGCCCAAGCAGCAGCAACCTGAAAGACCTGCAGGACTTA 420
QY 421 CACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACCTTCCAG 480
Db 1017 CACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGGAACCTTCCAG 480
QY 481 CTCCTTTGAGACGGATATTGTGAACATCATTAGCAATATCAGTTACACGCCACACCTG 540
Db 1077 CTCCTTTGAGACGGATATTGTGAACATCATTAGCAATATCAGTTACACGCCACACCTG 540
QY 541 CGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGACACAGATACCCCTTACCAA 600
Db 1137 CGGACGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGACACAGATACCCCTTACCAA 600
QY 601 CACACAGATGATCGACCTCTTGAATAATACCTGGCCCAACATCGGTTTGGATTCTGTT 660
Db 1197 CACACAGATGATCGACCTCTTGAATAATACCTGGCCCAACATCGGTTTGGATTCTGTT 660
QY 661 TCTCTCAGATGCAACAAGATTGATGAGGTGAGGTTAGACACTGAGTCAAGTCAAGAAAT 720
Db 1257 TCTCTCAGATGCAACAAGATTGATGAGGTGAGGTTAGACACTGAGTCAAGTCAAGAAAT 720
QY 721 TCAGTGATTATGAAGAAATGAAGCTAGTAGCTCCCAAGCATGGTCAAGTCAAGAAAT 780
Db 1317 TCAGTGATTATGAAGAAATGAAGCTAGTAGCTCCCAAGCATGGTCAAGTCAAGAAAT 780
QY 781 TTTTACAATACTACAAGGTCCACCGGGCCCGAGGGTCCCAAGAGGTGACAGAGGATCCCAAG 840
Db 1377 TTTTACAATACTACAAGGTCCACCGGGCCCGAGGGTCCCAAGAGGTGACAGAGGATCCCAAG 840
QY 841 GGACCCCTTGGCCCACTGGCAACAAGAGGACAGAAAGAGAGAGAGGGGAGGCTGGACCA 900
Db 1437 GGACCCCTTGGCCCACTGGCAACAAGAGGACAGAAAGAGAGAGAGGGGAGGCTGGACCA 900
QY 901 CCTGGCCCTGCGGGTGAGAGAGGCCCAATTGGACCAAGCTGGTCCCCCGGAGAGCGTGGC 960
Db 1497 CCTGGCCCTGCGGGTGAGAGAGGCCCAATTGGACCAAGCTGGTCCCCCGGAGAGCGTGGC 960
```

RESULT 2

```
US-09-954-456-782
; Sequence 782, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
```



```

; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..475
; NAME/KEY: CDS
; LOCATION: 476..964
; NAME/KEY: 3'UTR
; LOCATION: 965..1133
; NAME/KEY: polyA_signal
; LOCATION: 1101..1106
; NAME/KEY: polyA_site
; LOCATION: 1118..1133
; US-09-924-340-57

Query Match          6.5%; Score 110.6; DB 9; Length 1133;
Best Local Similarity 52.9%; Pred. No. 5.6e-20;
Matches 267; Conservative 0; Mismatches 229; Indels 9; Gaps 1;

QY 759 GCATGTCAGCTCATCAAGAATTTTACAATACTACAAAGTCCACCGGGCCCCCAGGGGTCC 818
Db 109 GCAAGAGCAGGGCCCCCAGGATCTGAAGGCTTCCAGGCCCCCGCCAGCGGGTCC 168
QY 819 AAGAGGTGACAGAGGATCCAGGACCCCTGCGCCCAACTGGCAACAAGGGACAGAAAGG 878
Db 169 CAGAGGAGCGAGGACCCCAAGGTAACCTCGGTGAGAAGGGCGACAGGATTTTCAAGG 228
QY 879 AGAAGAGGGGGAGCCCTGGACCACTTGGCCCTCGGGGTGAGAGAGGCCCAATTGGACCA 938
Db 229 CCAGCCAGGCTTTCCGGGCCACCGGGTCCCTGGATTCCCGAGGCAAAAGTTGGATCACC 288
QY 939 TGGTCCCCCGG-----AGAGCGTGGCGCAAGAGATCTAAAGGCTCCAGGGGCC 989
Db 289 TGGCCACCTGGCCCTCAAGCAGAGAGAGGCGACGAAGGATTCGAGGCCCATCAGGCCT 348
QY 990 CAAAGGCTCCGTGGTTCCTTGGGAAGCCCGGCCCTCAGGGCCCCCGCCCTCGGTCATCAGG 1049
Db 349 GCCTGGCTCCCTTGGGCCACCGGGACCTCTCTGGGATTCAGGGCCCCCGCGGCTTGGATGG 408
QY 1050 CCCCCGGGCCACACAGGCAAGAGGGACTCCCGGCCCTCAGGGCCCTCCTGGCTTCCA 1109
Db 409 TTTGGATGGGAAGGATGGCAAGCTGGCTTGAGGGGGGACCTTGGCTCTGCTGGCCCCC 468
QY 1110 GGGACTTCAGGACCCGTTGGGGAGCCTGGGGTGCCTGGACCTCGGGGACTGCCAGGCTT 1169
Db 469 TGGACTCATGGGACCAACCGGCTTTAAGGGAAACAGGACATCTCTGGCTCCAGGACC 528
QY 1170 GCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCAGGGCCCCCTCTGGCCCATCAGG 1229
Db 529 TAAGGCTGACTTGGCAAAACAGGTCCTCTGGCAGCACTGGCGGCCCTGGCGCAGAGG 588
QY 1230 AGCGGTGGTGGCCCTGGCCCTGCAG 1254
Db 589 TGAACCTGGTGCATGGGACCCAG 613

RESULT 5
US-09-992-600A-57
; Publication 57, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephanie
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.USA.DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29

```

```

; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: JPatent
; SEQ ID NO 57
; LENGTH: 1133
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..475
; NAME/KEY: CDS
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 965..1133
; NAME/KEY: polyA_signal
; LOCATION: 1101..1106
; NAME/KEY: polyA_site
; LOCATION: 1118..1133
; US-09-992-600A-57

Query Match          6.5%; Score 110.6; DB 9; Length 1133;
Best Local Similarity 52.9%; Pred. No. 5.6e-20;
Matches 267; Conservative 0; Mismatches 229; Indels 9; Gaps 1;

QY 759 GCATGTCAGCTCATCAAGAATTTTACAATACTACAAAGTCCACCGGGCCCCCAGGGGTCC 818
Db 109 GCAAGAGCAGGGCCCCCAGGATCTGAAGGCTTCCAGGCCCCCGCCAGCGGGTCC 168
QY 819 AAGAGGTGACAGAGTATCCAGGACCCCTGCGCCCAACTGGCAACAAGGGACAGAAAGG 878
Db 169 CAGAGGAGCGAGGACCCCAAGGTAACCTCGGTGAGAAGGGCGACAGGATTTTCAAGG 228
QY 879 AGAAGAGGGGGAGCCCTGGACCACTTGGCCCTCGGGGTGAGAGAGGCCCAATTGGACCA 938
Db 229 CCAGCCAGGCTTTCCGGGCCACCGGGTCCCTGGATTCCCGAGGCAAAAGTTGGATCACC 288
QY 939 TGGTCCCCCGG-----AGAGCGTGGCGCAAGAGATCTAAAGGCTCCAGGGGCC 989
Db 289 TGGCCACCTGGCCCTCAAGCAGAGAGAGGCGACGAAGGATTCGAGGCCCATCAGGCCT 348
QY 990 CAAAGGCTCCGTGGTTCCTTGGGAAGCCCGGCCCTCAGGGCCCCCGCCCTCGGTCATCAGG 1049
Db 349 GCCTGGCTCCCTTGGGCCACCGGGACCTCTCTGGGATTCAGGGCCCCCGCGGCTTGGATGG 408
QY 1050 CCCCCGGGCCACACAGGCAAGAGGGACTCCCGGCCCTCAGGGCCCTCCTGGCTTCCA 1109
Db 409 TTTGGATGGGAAGGATGGCAAGCTGGCTTGAGGGGGGACCTTGGCTCTGCTGGCCCCC 468
QY 1110 GGGACTTCAGGACCCGTTGGGGAGCCTGGGGTGCCTGGACCTCGGGGACTGCCAGGCTT 1169
Db 469 TGGACTCATGGGACCAACCGGCTTTAAGGGAAACAGGACATCTCTGGCTCCAGGACC 528
QY 1170 GCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCAGGGCCCCCTCTGGCCCATCAGG 1229
Db 529 TAAGGCTGACTTGGCAAAACAGGTCCTCTGGCAGCACTGGCGGCCCTGGCGCAGAGG 588
QY 1230 AGCGGTGGTGGCCCTGGCCCTGCAG 1254
Db 589 TGAACCTGGTGCATGGGACCCAG 613

RESULT 6
US-09-925-302-64
; Sequence 64, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 2691
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2653)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2667)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2683)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-925-302-64

Query Match 6.5%; Score 109.4; DB 10; Length 2691;
Best Local Similarity 54.4%; Pred. No. 2e-19;
Matches 240; Conservative 1; Mismatches 197; Indels 3; Gaps 1;

QY 793 CAAGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCAGAGACCCCGCTGCG 852
Db 31 CCAGGAGCACCAGGTGTCCAGGGCCCAAGGAGACCCCTGGATTCCAGGGCATGCTGT 90

QY 853 CCAACTGGCAACAGGGACAGAAAGAGAGAGAGGGGAGCCTGGACACCTGGCCCTGGG 912
Db 91 ATTGTGTGCTCTCCAGGAATCACAGGCTTAAGGGTGATATGGGCGCTCCAGGAGTTCCA 150

QY 913 GTGTAGAGAGGGCCCAATTTGACACAGCTGTGTCCCGCCCGGAGAGCGTGGGCAAGAGTCT 972
Db 151 GGAATTCAGGTCCAAAGGTCTTCTGTGCTCCAGGAGTAATTAAGGTGATCAAGCGCAT 210

QY 973 AAAGGCTCCAGGGCCCGCAAGGCTCCCGTGTTCCTTGGGAAGCCCGGCCCTCAGG-- 1030
Db 211 CAMGGCGTCCCGGAGCTAAAGGTCTCCCGGTCTCTCGGCTCCCGCCCGCAGGTCTTACGAC 270

QY 1031 -CCCCAGTGGGACCCAGGCCCCCGGGCCACACAGGCAAGAGAGGACTCCCGGCCCT 1089
Db 271 ATCATCAAGGGGAGCGGGGCTCCCTGTGTGAGGGCCCCCGGCGTGAAGGGGCTT 330

QY 1090 CAGGGCCCTCTCTGCTTCCAGGAGCTTCAGGGACCGTGTGGGAGCGCTGGGTGCTGGA 1149
Db 331 CAGGAGCTCCAGGGCCCGAAGGCCAGCAAGGTGTACAGGATTTGGTGTATACCTGGA 390

QY 1150 CTTGGGGAGTCCAGGCTTGCTGGGTAGCAGCATGCCAGGCCCCCAAGGGCCCCCCC 1209
Db 391 CTTCCAGGTATCTCTGGGTTTACGGTGTCCCTGTGCGCCAGAAAGAGAGATGGACCTGCC 450

QY 1210 GGGCTCTCTGGCCCATCAGGA 1230
Db 451 GGGCTACTGTGTCAGAGGA 471

RESULT 7
US-09-925-301-42
; Sequence 42, Application US/09925301
; Patent No. US2002052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301

; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 2192
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-301-42

Query Match 6.3%; Score 107.2; DB 10; Length 2192;
Best Local Similarity 53.5%; Pred. No. 7e-19;
Matches 274; Conservative 0; Mismatches 228; Indels 10; Gaps 2;

QY 796 GTTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCAGGAGACCCCTTGCCCA 855
Db 638 GGCCCCGCTGGACCCCGTGTCTCTCTGTGCTCTGTGCTGCTGCTGCTGCTGCTGCT 697

QY 856 ACTGGCAACAGGGACAGAAAGAGAGAGGGGAGCCTTGGACACCTTGCCCTGGCGGT 915
Db 698 GCTGCAAGAGTGTGTGCTGTGAGACTGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 757

QY 916 GAGAGAGGCCCAATTGGACAGCTGTGCTCCCGCCGAGAGCGTGGCGGCAAGAGATCTAAA 975
Db 758 CCGTTGGCGCGCTGGCCCGCCGACCCCAAGGCCCGCTGGTGACAAGGGTGAGACA 817

QY 976 GGTCTCCAGGGCCCGCAAGGCTCCCGTGTTCCTTGGGAAGCCCGGCCCTCAGGGCCCC 1035
Db 818 GCGCAACAGGGGACAGAGGATAAAGGTACACCTGGCTTCTGTGCTGCTGCTGCTGCTG 876

QY 1036 AGTGGGGACCCAGGCCCCCGGGCCACACAGCAAGAGAGGACTCCCGGCCCTCAGGGC 1095
Db 877 -----CCCTGGGCCCTCTCTGTGTGACAAGGTCCCTCTGTGAGCCTCTGTT 928

QY 1096 CTTCTGTGCTTCCAGGAGCTTCAGGGCACCGTTGGGGAGCCTTGGGTGCTTGACCTCGG 1155
Db 929 CCGTGTGCTGCTCCAGGTCCTCCCTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 988

QY 1156 GCACTGCCAGGCTTGTGCTGGGTACAGGCATGCCAGGGCCCCAAGGGCCCCCGGCCCT 1215
Db 989 GGTCTCCCTGGGCCCATTTGGGCCCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1048

QY 1216 CTTGCCCCATCAGGAGCGG-TGGTGCCCTGCGCTTGCAGATGAGCCAAACCCCGCAC 1274
Db 1049 GTTGTGCCCCCGGCCCTCTGTGACCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1108

QY 1275 GGAGGACATGCTGCCCGCTCCTACTGGAAGA 1306
Db 1109 GACTTCAGCTTCTGCCCCAGCCACCTCAAGA 1140

RESULT 8
US-09-880-107-3947
; Sequence 3947, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1

```

; SEQ ID NO 3947
; LENGTH: 5086
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274616
; NAME/KEY: unsure
; LOCATION: (1)..(5086)
; OTHER INFORMATION: n = a o r c o r g o r t
US-09-880-107-3947

Query Match          6.3%; Score 107; DB 10; Length 5086;
Best Local Similarity 51.3%; Pred. No. 1.3e-18;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

QY 789 ACTACAGGTCCACCGGGCCCGAGGGTCCAAAGAGTGACAGAGTCCACAGGGACCC 848
DB 709 ACAGCCCGTGTCTGGTGTGAAGGGTGAACCTGGTGGCCCTGGTGAATAATGGAAC 768
QY 849 TGGCCCACTGGCAACAAGGGACAGAAAGAGAGAGGGGAGCCTGGACCACTGGCC 908
DB 769 AGGTCAACAGAGCCGTGGCTTCTGGTGAGAGAGACGTGTGGTCCCTGGCC 828
QY 909 TCGGGGTGAGAGGCCCAATTTGACCACTGGTCCCCCGGAGAGCGTGGCGGCAAGG 968
DB 829 AGCTGTGCGCGTGGCAGTGATGGAAGTGTGGGTCCCGTGGGTCTGCTGGTCCCAT 888
QY 969 ATCTAAGGCTCCACGGGCCCAAGGCTCCGTTGGTTCCTTGGAGAGCCCGCCCTCA 1028
DB 889 GTCTGTGCGCCCTCCAGGCTTCCAGGTCCTCCAGGTCCTGGCCCAAGGGTGAAT 948
QY 1029 GGGCCCCAGTGGGAGCCACAGGCCCGCCGACAGCAAGAGGGACTCCCGGGCC 1088
DB 949 TGGTACGCTGTGCTGTGCTGGTCCCGCGCTCCCGTGGTGAAGTGGGTCTCCAGGC 1008
QY 1089 TCAGGCCCTCTGGCTTCCAGGAGCTTCAGGACACCGTTGGGAGCGCTGGGGTGG 1148
DB 1009 CTCGGCCCCGTTGGACCTTCTGGTAATCTTGAGCAACAGCCCTTACTTGGTGC 1068
QY 1149 ACTCGGGAGCTGCCAGGCTTGCCTGGGTACAGGATGCCAGGCCCAAGGCCCGCC 1208
DB 1069 TGTGTGCGCTTCCCGCGTGTGTGGGCTCCCGGCTTCCAGCCCTGGAGCGGTATCC 1128
QY 1209 CGGCCCTCTGCGCCATCAGGAGCGGTGTGCGCCCTGGCCCTGCAGAAATGAGCC 1268
DB 1129 TGGCCCTGTGTGTGCTGCGGCTGCTACTGCTCCAGAGGACTTGTGGTGGCTGG 1188
QY 1269 GGC 1271
DB 1189 AGC 1191

RESULT 9
US-09-925-299-206
; Sequence 206, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Query Match          6.3%; Score 107; DB 9; Length 5145;
Best Local Similarity 51.3%; Pred. No. 1.3e-18;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

QY 789 ACTACAGGTCCACCGGGCCCGAGGGTCCAAAGAGTGACAGAGTCCACAGGGACCC 848
DB 712 ACAGCCCGTGTCTGGTGTGAAGGGTGAACCTGGTGGCCCTGGTGAATAATGGAAC 771
QY 849 TGGCCCACTGGCAACAAGGGACAGAAAGAGAGAGGGGAGCCTGGACCACTGGCC 908
DB 772 AGGTCAACAGGAGGCCGTGGCTTCTGGTGAGAGAGGACGTGTGGTGGCCCTGGCC 831
QY 909 TCGGGGTGAGAGAGGCCCAATTTGGACCACTGGTCCCCCGGAGAGCGTGGCGGCAAGG 968
DB 832 AGCTGTGCGCTGGCAGTGATGGAAGTGTGGGTCCCGTGGGTCTGCTGGTCCCAT 891
QY 969 ATCTAAGGCTCCACGGGCCCAAGGCTCCGTTGGTTCCTTGGGAAGCCCGGCCCTCA 1028
DB 892 GTCTGTGCGCCCTCCAGGCTTCCAGGTCCTCCAGGTCCTGGCCCAAGGGTGAAT 951
QY 1029 GGGCCCCAGTGGGAGCCCGAGGCCCGCCGACAGCAAGAGGGACTCCCGGGCC 1088
DB 952 TGGTAACGCTGCTGTGCTGGTGGTCCCGCGCTCCCGTGGTGAAGTGGGTCTTCCAG 1011
QY 1089 TCAGGCCCTCTGCGCTTCCAGGAGCTTCCAGGGCACTTCCAGGGAGCCTTGGGGTGG 1148
DB 1012 CTCGGCCCCGTTGGACCTCTCTGGTAATCTTGAGCAACAGCCCTTACTTGGTG 1071
QY 1149 ACTCGGGAGCTGCCAGGCTTGCCTGGGTACAGGATGCCAGGCCCAAGGGCCCGCC 1208
DB 1072 TGTGTGCTGGCTTCCCGGCTTGGTGGGCTCCCGGCTTCCCGTGGACCCCGGGTAT 1131
QY 1209 CGGCCCTCTGCGCCATCAGGAGCGGTGTGCGCCCTGGCCCTGCAGAAATGAGCC 1268
DB 1132 TGGCCCTGTGTGTGCTGGCGGTGCTACTGCTGCAGAGGACTTGTGGTGGCTGG 1191
QY 1269 GGC 1271
DB 1192 AGC 1194

RESULT 10
US-09-925-299-206
; Sequence 206, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145

```

```
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.13
US-10-044-090-22

Query Match      6.3%; Score 107; DB 12: Length 5432;
Best Local Similarity 51.3%; Pred. No. 1.3e-18;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps

QY   789 ACTTAAAGGTCCACCGGGCCCCAGGGTCCAAGAGGTCACAGAGATCCCAGGACNCCCC 848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1042 ACAGCCCGTGCTCTCTGTGTGAAGGGTGAACCTGGTGCCTCGTTGAAATGGAATCC 1101

QY   849 TGCCCAACTGGCAACAAGGGACAGAAAAGGAGAGAAGGGAGCCTGGACCACCTGGGCC 908
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1102 AGGTCAACAGGAGCCCGTGGGCTTCCTGGTGAGAGAGGACGTGTTGGTGCCTTGCCC 1161

QY   909 TGGCGGTGAGAGAGGCCCAATTGGACAGCTGGTCCCCCGGAGAGAGCCTGGCGGCAAGG 969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1162 AGCTGGTGGCCGTGGCAGTGATGGAAGTGTGGGTCCCGTGGGTCCCTGCTGCTCCATTGG 1221

QY   969 ATCTAAGGCTCCACAGSGCCCCAAAGGCTCCCGTGGTTCCTCGGAAAGCCCGCCCTCA 1028
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1222 GTCTGTGCCCCCTCCAGGCTTCCCAGGTGCCCTCGCCCCAAGGGTGAATTTGAGCTGT 1281

QY   1029 GGSCCCCACTGGGGACCCACAGGCCCCCGGGGCCACACGAGCAAAGAGGACTCCCGGCC 1088
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1282 TGATAAGCTGGTCTCTGTGTGCCCGCGGTCCCGTGGTGAAGTGGGTCTTCCAGCGCT 1341

QY   1089 TCAGGCGCCTCTCTGGCTTCCAGGGACTTCAGAGGCACCGTTGGGAGCCTGGGGTGCCCTGG 1148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1342 CTCGGGCCCGCTGGACCTCTCTGGTAATCTCGAGCAAAAGGCGCTTACTGCTGCAAGGG 1401

QY   1149 ACCTCGGGAGCTCCAGGCTTGCCTGGGTACAGGCATGCCAGGCCCCAGAGGGCCCCC 1208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1402 TGCTGTGCCCCTTCCCGGGGTTGCTGGGCTCCCGGCTCCCTGGACCCCGCGGTATTCC 1461

QY   1209 CGGCGCTCTCGGCCCATCAGGAGCGTGTGTCCTCGCCCTGCGCCTGCAAGTAGCAACCCC 1268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   1462 TGGCCCTGTGGTCGCGGTGCTACTGGTCCAGAGGACTGTTGGTGAGCCCTGGTGCC 1521

QY   1269 GGC 1271
    ||
Db   1522 AGC 1524

RESULT 12
US-10-001-887-33
; Sequence 33, Application us/10001887
; Patent No. US2002015464A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; FILE REFERENCE: DEX-0269
; CURRENT APPLICATION NUMBER: US/10/001,887
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,998
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,563
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 4908
; TYPE: DNA
; ORGANISM: Homo sapien
; US-10-001-887-33
```



```

Db 957 AGGGCTAGCGGTGTGCCAGGCGCCAGCTTGGGACAAAAGAGGAGGCGCTTGGAGACCAAGGTG 1016
QY 962 GCAAGAGTCTAAAGGCTCCAGGCGCCCAAGAGGTCCCGTGTTCCTCGGAAAGCCCG 1021
Db 1017 AGCGGGCCCGCAGGCGCTTCTGGATCTCTGTGTCCTCGGAAAGAGGAGAGCCAG 1076
QY 1022 GCGCTCAGGGCCCCAGTGGGAGCCAGGCGCCCGGCGGCCACAGGCAAGAGGAGCTCC 1081
Db 1077 GCGCTCAGGAGAAATTTGTCGCCAGGCGATCATGGACAGAGGAGTACCAAGGCGAGA 1136
QY 1082 CCGGCGCTCAGGCGCTCTGTGCTTCCAGGAGCTTCAGGCGACCGTGGGAGCGCTGGG 1141
Db 1137 GGGGTCCAGTGGGCAACAGGCGCTTCAGGGAAGCAGGCGCTTAAGGGGAGCAGGCGCC 1196
QY 1142 TGCCTGCGACTCGGAGCTGCCAGGCTTGCCTGGGGTACAGGCGATGCCAGGCGCCCAAGG 1201
Db 1197 CCGCGGATTCAGGCGCCCAAGGCTTGCAGGCGTTCAGGAGGAGCAAGGCTCCCGAG 1256
QY 1202 GCGCGCGCGCGCTCCTGCGCCCATCAGGAG 1231
Db 1257 GGAAGACCGGCGCGCGGCAAGTGGGTG 1286

```

RESULT 15

US-09-954-456-786
; Sequence 786, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:

; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand

; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 786

; LENGTH: 5416

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-954-456-786

Query Match 6.2%; Score 105.4; DB 10; Length 5416;
Best Local Similarity 51.1%; Pred. No. 3.6e-18;
Matches 247; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

```

QY 789 ACTAAGGTCCACCGGCGCCAGGCGTCCAAAGAGGTGCACAGAGGATCCAGGAGACCC 848
Db 1039 ACAGCCCGGTGCTCTGTGTGAAGGTGAACCTGGTGCCTCGTGAAGTGAAGTGAAC 1098
QY 849 TGGCCCACTGCACAGGAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 908

```

```

Db 1099 AGGTCAACACAGAGAGCCCGTGGTCTTCTCGTGAGAGAGACGCTGTGTGGTCCCGCTGGTCC 1158
QY 909 TCGGGGTGAGAGAGAGGCGCCCAATTTGGACCACTGTCTCCCGCGGAGAGCGCTGGCGGCAAGG 968
Db 1159 ACCTGGTGGCCCTGGAAAGTGATGGAAGTGGTGGTCCCGTAGGTCTGCTGCTTAAATGG 1218
QY 969 ATCTAAAGGCTCCCAAGGCGCCCAAGAGTCCCGTGGTTCCTCGTGGAAAGCGCGGCCCTCA 1028
Db 1219 GTCTGTGGCCCTCCAGGTTTCCAGGTCCCGTGGTCCCAAGGCTGAAATTTGGAGCTGT 1278
QY 1029 GGGCCCCAGTGGGAGCCAGGCGCCCGGCGCCACAGGCAAGAGGAGGACTCCCGGCGCC 1088
Db 1279 TGGTAACGCTGTCTTACTTGGACCCCGCGTCCCGTGGTGAAGTGGGTCTTCCAGGCTT 1338
QY 1089 TCAGGCGCTCTCTGGCTTCCAGGAGCTTCAGGCGCACCGTGGGAGAGCTTGGGTGCTGG 1148
Db 1339 CTCGCGCCCGTGGACCTCTCTGGTAATCTCGAGCAACAGCGCTTACTGGTCCCAAGG 1398
QY 1149 ACCTCGGGAGCTGCGAGGCTTGCCTGGGGTACAGGAGTGCAGGCGCCCAAGGCGGCC 1208
Db 1399 TGCTGTGGCTTCCCGCGTTCCTGGGCTCCCGGCTCCCGGCTCCCGTGGACCCCGGCTATTC 1458
QY 1209 CGGCGCTCTGCGCCATCAGAGCGGTGGTGGCCCTGGCGCTGCAGATGAGCCCAACCC 1268
Db 1459 TGGCGCTCTGTGTGCTGCGGTACTACTGTGTCAGAGGAGCTTGTGTGGTGAAGCTTGGTCC 1518
QY 1269 GGC 1271
Db 1519 AGC 1521

```

Search completed: March 21, 2003, 08:22:02
Job time : 277.28 secs

Gencore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:34:06 ; Search time 2223.7 Seconds
(without alignments)

12344.923 Million cell updates/sec

Title: US-09-763-712a-1_COPY_1_1695

Perfect score: 1695

Sequence: 1 gtccacgaatcgcagcaaga.....cagtaactgcatcgtcatta 1695

Scoring table: IDENTITY_NUC

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 1357.4 | 80.1 | 3305 | 11 | BC009162 Mus muscu |
| 2 | 677.8 | 40.0 | 957 | 14 | B0955927 AGENCOURT |
| 3 | 656.4 | 38.8 | 936 | 14 | B0891432 AGENCOURT |
| 4 | 616.8 | 36.4 | 884 | 14 | B0934501 AGENCOURT |
| 5 | 559 | 33.0 | 906 | 13 | B1456109 AGENCOURT |
| 6 | 556.4 | 32.8 | 861 | 14 | B0713873 AGENCOURT |

| | | | | | | |
|---|-------|-------|------|------|----------|----------|
| 7 | 525.4 | 31.0 | 668 | 14 | B0674807 | |
| c | 8 | 488.6 | 28.8 | 511 | 12 | BF805750 |
| c | 9 | 480.2 | 28.3 | 581 | 14 | B0127513 |
| c | 10 | 475.4 | 28.0 | 967 | 9 | AL568743 |
| c | 11 | 445.6 | 26.3 | 580 | 12 | BE910803 |
| c | 12 | 437.2 | 25.8 | 552 | 10 | AN958053 |
| c | 13 | 415.8 | 24.5 | 601 | 10 | BE290299 |
| c | 14 | 395.8 | 23.4 | 410 | 12 | BF856523 |
| c | 15 | 369.4 | 21.8 | 638 | 10 | BB248064 |
| c | 16 | 351.4 | 20.7 | 500 | 14 | BM676508 |
| c | 17 | 350 | 20.6 | 507 | 14 | BM713891 |
| c | 18 | 343.8 | 20.3 | 808 | 14 | B0771366 |
| c | 19 | 311 | 18.3 | 339 | 9 | AA304251 |
| c | 20 | 301.4 | 17.8 | 357 | 10 | AM379436 |
| c | 21 | 297.2 | 17.5 | 517 | 9 | AL543000 |
| c | 22 | 296 | 17.4 | 310 | 9 | AA361740 |
| c | 23 | 295.2 | 17.5 | 679 | 14 | W27610 |
| c | 24 | 271 | 16.0 | 350 | 9 | AA012704 |
| c | 25 | 258.2 | 15.2 | 286 | 12 | BF854324 |
| c | 26 | 245.8 | 14.5 | 420 | 12 | BE929813 |
| c | 27 | 244.2 | 14.4 | 252 | 13 | BT051416 |
| c | 28 | 244 | 14.4 | 277 | 12 | BF854320 |
| c | 29 | 242.2 | 14.3 | 382 | 12 | BG012097 |
| c | 30 | 241 | 14.2 | 550 | 12 | BF858081 |
| c | 31 | 233.8 | 13.8 | 429 | 12 | BE929798 |
| c | 32 | 230.8 | 13.6 | 465 | 10 | AM240221 |
| c | 33 | 227.6 | 13.4 | 930 | 9 | AL541116 |
| c | 34 | 225 | 13.3 | 922 | 9 | AL542672 |
| c | 35 | 223.2 | 13.2 | 280 | 13 | BT010883 |
| c | 36 | 212.6 | 12.5 | 281 | 12 | BF854408 |
| c | 37 | 212 | 12.5 | 723 | 14 | BO004176 |
| c | 38 | 210.2 | 12.4 | 943 | 14 | BO887163 |
| c | 39 | 209.8 | 12.4 | 628 | 14 | BM721855 |
| c | 40 | 205.8 | 12.1 | 261 | 13 | BT010880 |
| c | 41 | 205.6 | 12.1 | 1082 | 14 | BM907108 |
| c | 42 | 201 | 11.9 | 655 | 12 | BG482931 |
| c | 43 | 200.8 | 11.8 | 345 | 9 | AT742661 |
| c | 44 | 199.6 | 11.8 | 441 | 10 | BF749743 |
| c | 45 | 193.4 | 11.4 | 234 | 12 | BF854188 |

ALIGNMENTS

| | | | | | |
|------------|--|---------|--------------|--------------|--|
| RESULT 1 | BC009162 | 3305 bp | mus musculus | mus musculus | Similar to scavenger receptor with C-type lectin, clone IMAGE:2811487, mRNA. |
| LOCUS | BC009162 | | | | |
| DEFINITION | BC009162 | | | | |
| ACCESSION | BC009162 | | | | |
| VERSION | BC009162.1 | | | | GI:14714370 |
| KEYWORDS | HTC. | | | | |
| SOURCE | house mouse. | | | | |
| ORGANISM | Mus musculus | | | | |
| REFERENCE | 1 (bases 1 to 3305) | | | | |
| AUTHORS | Strausberg, R. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA | | | | |
| REMARK | NIH-MGC Project URL: http://mgc.nci.nih.gov | | | | |
| COMMENT | Contact: MGC help desk Email: cgaabs-r@mail.nih.gov Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ | | | | |

Contact: amg@bcm.tmc.edu
Gunnar, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/BLINL at: <http://image.lnl.gov>
Series: IRAC Plate: 5 Row: 5 Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis

This clone has the following problem: incomplete processing.

FEATURES

Location/Qualifiers

Source

1..3305
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="IMAGE:2811487"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_id="NCI_CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: PCMV-SpOrf6"

BASE COUNT 1031 a 757 c 760 g 757 t

ORIGIN

Query Match 80.1%; Score 1357.4; DB 11; Length 3305;

Best Local Similarity 87.6%; Pred. No. 0;

Matches 1484; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

QY 1 GTCCGATATGCGACGATACCGAGCGTCTCCAGGCAATTCGAGAACCAATGTAT 60
DB 627 GTCCGATATGCGACGATACCGAGCGTCTCCAGGCAATTCGAGAACCAATGTAT 686
QY 61 TCTCATATGTCGATCATGAACTCAACACCTGAACCTGACCCAGGCGTACAGAG 120
DB 687 TCTCATATGTCGATCATGAACTCAACACCTGACCCAGGCGTACAGAG 746
QY 121 AACCTCATGAGATTCGACGCGTCTGTGATGACACAAAGCCAGCTATCCAGCATC 180
DB 747 AACCTCATGAGATTCGACGCGTCTGTGATGACACAAAGCCAGCTATCCAGCATC 806
QY 181 AAGACGACTTTTCAAAATTCGACGAGTTTCTTCAAGCCAAAGAGACGAGATTGG 240
DB 807 AAGACGACTTTTCAAAATTCGACGAGTTTCTTCAAGCCAAAGAGACGAGATTGG 866
QY 241 CTGAAGAGAAAGTCGAGAGCTTCGACAGCGTGGCTCCAAACACTCTGGGTTGGCCAAA 300
DB 867 CTGAAGAGAAAGTCGAGAGCTTCGACAGCGTGGCTCCAAACACTCTGGGTTGGCCAAA 926
QY 301 GCCAACAAGACACCTCGAGAGATATGACAGCCAGCTCAACTCATTCACAGGTCAAGTG 360
DB 927 GCCAACAAGACACCTCGAGAGATATGACAGCCAGCTCAACTCATTCACAGGTCAAGTG 986
QY 361 GAGACATTCACATATCTCTCAAGCCAGAGAGCAACTGAAAGCTTCGAGAGATTGA 420
DB 987 GAGACATTCACATATCTCTCAAGCCAGAGAGCAACTGAAAGCTTCGAGAGATTGA 1046
QY 421 CACAAAGATGAGAGATGAGACAGCCATCAAGTTCAACCACTGAGGAGGAGGTTTCAG 480
DB 1047 CACAAAGATGAGAGATGAGACAGCCATCAAGTTCAACCACTGAGGAGGAGGTTTCAG 1106
QY 481 CTCTTTGAGAGGATATGTAACATCATTAAGCAATATCAAGTTACACAGCCCAACCTG 540
DB 1107 CTCTTTGAGAGGATATGTAACATCATTAAGCAATATCAAGTTACACAGCCCAACCTG 1166
QY 541 CGGACGCTGACGAGCAATTAATGAAGACGACCTTCGACAGATACCTTACCAAA 600
DB 1167 CGGACGCTGACGAGCAATTAATGAAGACGACCTTCGACAGATACCTTACCAAA 1226
QY 601 CACACAGATGATGACCTCTGTAATATACCTGCGCAAACTCGTTGGATTCGTT 660
DB 1227 CACAGATGATGACCTCTGTAATATACCTGCGCAAACTCGTTGGATTCGTT 1286

QY 661 TCTCTCAGAGATGCAACAGATTTTGATAGGTGAGATGACACTGGAAGCCCACTTA 720
DB 1287 TCTCTCAGAGATGCAACAGATTTTGATAGGTGAGATGACACTGGAAGCCCACTTA 1346
QY 721 TCACTGATTTAGAGAAATGAAAGCTAGTACCTCCAGCATGGTCACTATCAAGAT 780
DB 1347 TCACTGATTTAGAGAAATGAAAGCTAGTACCTCCAGCATGGTCACTATCAAGAT 1406
QY 781 TTTCAATTAATACAGAGTCCACCGGCGCCAGAGGCTCCAAAGAGTGCAGAGATCCAG 840
DB 1407 TTTCAATTAATACAGAGTCCACCGGCGCCAGAGGCTCCAAAGAGTGCAGAGATCCAG 1466
QY 841 GGACCCCTGAGCCCACTGGCAACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 1467 GGACCCCTGAGCCCACTGGCAACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1526
QY 901 CTTGGCCCTGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 1527 CTTGGCCCTGCGGCTGAG 1586
QY 961 GGCAGAGATCTAAAGGCTCCAGAGGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 1587 GGCAGAGATCTAAAGGCTCCAGAGGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1646
QY 1021 GGCCCTGAGGCGCCCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 1647 GGCCCTGAGGCGCCCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1706
QY 1081 CCGGCGCCCTGAGGCGCCCTCTGCTGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1707 CCGGCGCCCTGAGGCGCCCTCTGCTGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1766
QY 1141 GTGCTGAGCTGAGGAG 1200
DB 1767 GTGCTGAGCTGAGGAG 1826
QY 1201 GGCCCGCCCGGCGCCCTCTGCTGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1827 GGCCCGCCCGGCGCCCTCTGCTGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1886
QY 1261 CCAACCCCGGCGCGGAG 1320
DB 1887 CCAACCCCGGCGCGGAG 1946
QY 1321 TGTCTACTATTTTTCAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1947 TGTCTACTATTTTTCAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2006
QY 1381 AAGTCTTACATCTTGTTCATTAACACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 2007 AAGTCTTACATCTTGTTCATTAACACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2066
QY 1441 ATGCTAGGAG 1500
DB 2067 ATGCTAGGAG 2126
QY 1501 AAGTGGCTGAGTGGAG 1560
DB 2127 AAGTGGCTGAGTGGAG 2186
QY 1561 TGGGGTATGAGGAG 1620
DB 2187 TGGGGTATGAGGAG 2246
QY 1621 AAGGATTTTCAATGAG 1680
DB 2247 AAGGATTTTCAATGAG 2306
QY 1681 CTGTCATCTGCAATTA 1695
DB 2307 CTGTCATCTGCAATTA 2321

```

RESULT 2
LOCUS      B0955927              957 bp      mRNA      linear      EST 21-AUG-2002
DEFINITION AGENCOURT.8763247 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6313500
5', mRNA sequence.
ACCESSION  B0955927
VERSION    B0955927.1 GI:22371405
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 957)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cga@bbs-remail.nih.gov
            Tissue Procurement: Susan L. Sullivan, Ph.D.
            cDNA Library Preparation: Resgen, Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13740 row: 1 column: 13
            High quality sequence stop: 665.
FEATURES
    source
        1..957
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone="IMAGE:6313500"
        /lab_host="NIH_MGC_129"
        /note="Organ: Olfactory epithelium; Vector:
        pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
        unidirectionally. Primer: Oligo dT. Average insert size
        2.2 kb. Constructed by: Resgen, Invitrogen Corp. Note: this
        is a NIH_MGC library."
BASE COUNT  260 a 269 c 258 g 169 t 1 others
ORIGIN
Query Match      40.0%; Score 677.8; DB 14; Length 957;
Best Local Similarity 86.9%; Pred. No. 2.3e-155;
Matches 780; Conservative 0; Mismatches 113; Indels 5; Gaps 3;
OY 330 CAGCAGCTCAATCATTCATGATGAGAGAAATCACCACATCTCTCAAGCCAA 389
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 13 CCGGATCTCAGCTCATTCACAGTCATGAGACAAATATACCATCTCTCAGAGCCAA 72
OY 390 CAGCAGAAACCTGAAAGACCTGCAGACTTACACAAAGTGCAGAGATGAAGACGCCAT 449
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 73 CGAGAGACCTCGAAAGACCTTCAGAACTTACACAGATACAGAAATAGAACACTCTG 132
OY 450 CAAGTTCACCACTGAGAGAGAGCTTCCAGCTTTGAGACGATATGTGAATCATCAT 509
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 133 CAAGTTCACCACTGAGAGAGAGCTTCCAGCTTTGAGACGATATGTGAATCATCAT 192
OY 510 TAGCAATATCACTTACACAGAGCCACCTGCGAGAGCTGACAGCAATCTAAATGAAGT 569
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 193 TAGCAATATCACTTACACAGAGCCACCTGCGAGAGCTGACAGCAATCTGATGATGT 252
OY 570 CAGGAGCACTTGCACAGATACCTTACCAACACACAGATGATGACTCTGATATGA 629
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 253 TAGGACCACTGACAGACAGCACTTACCAAGATGACCTGACCTCTTGAATATA 312
OY 630 TACCTTGCCACATCCGTTTGATTTCTCTCTCAGAGTACCAAGATTTGATGAG 689
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 313 CACACTAGTCAACATCCGCTTGATTTCTCTCTCAGAGTACCAAGATGATGAG 372
OY 690 GTGAGGTTAGCACTGAGATGACCACTTACAGTATGAGAAATGAAGCTAGT 749
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

```

DB 373 GTCAAGTTAGCACTGAGAGAGGAGCCCACTTATCAGTGTATGAGAGAGATGAATGCT 432
OY 750 AAGCTCCAGCATGTGTACACTCATCAAGATTTTAAATACATACAGATCCACGGGCC 809
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 433 TCACCTCAAGCAGCGGTACACTCATCAAGATTTTAAATACATACAGATCCACGGGCC 492
OY 810 CAGGGTCCCAAGAGGTGACAGAGATCCAGAGGAGCCCTGCGCCCACTGAGCAAGAGG 869
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 493 CAGAGGTCCCAAGAGGTGACAGAGATCTCAGGAGACCACTGTGTCAACTGACACAGAG 552
OY 870 ACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 929
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 553 ACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 612
OY 930 TGGACCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 989
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 613 TGGACCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 672
OY 990 CAAAGGCTCCGTTGTTCCCTGTGGAAGCCGCGCTCAGAGGAGAGAGAGAGAGAGAG 1049
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 673 CAAAGATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 732
OY 1050 CCCCCCGGGCCACACAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1109
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 733 ACCACAGGTCCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 791
OY 1110 GGGACTTCAGGGCAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1168
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 792 GGGACTTCAGGGCAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 851
OY 1169 TGGCTGGGTACAGCATATGC---AGGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1223
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 852 TGGCAGGGGTGCGAGCATATGCCCTGTGGGCTTAAGGAGACCACTGTGGGCTCCAGAGCC 909
RESULT 3
LOCUS      B0891432              936 bp      mRNA      linear      EST 16-AUG-2002
DEFINITION AGENCOURT.8754195 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6332754
5', mRNA sequence.
ACCESSION  B0891432
VERSION    B0891432.1 GI:22283446
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 936)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cga@bbs-remail.nih.gov
            Tissue Procurement: Mark Macdonochie, Ph.D. and Nancy L. Freeman,
            Ph.D.
            cDNA Library Preparation: Resgen, Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13790 row: n column: 19
            High quality sequence start: 25
            High quality sequence stop: 632.
FEATURES
    source
        1..936
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone="IMAGE:6332754"
        /lab_host="NIH_MGC_130"
        /note="Organ: Olfactory epithelium; Vector: pCMV-SPORT6.1.cdb;
        Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally."

```



```

DEFINITION  AGENCOURT_8482828 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6306240
ACCESSION   B0713873
VERSION     B0713873.1 GI:21852772
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE   NIH-MGC http://mgi.nci.nih.gov/
AUTHORS    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: Susan L. Sullivan, PhD.
            cDNA Library Preparation: Resgen, Invitrogen Corp.
            cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLM13721 row: n column: 01
            High quality sequence stop: 598.
FEATURES
    source
        1..861
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone="IMAGE:6306240"
        /clone_1ib="NIH_MGC_129"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: olfactory epithelium; Vector:
        pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
        unidirectionally. Primer: Oligo dt. Average insert size
        2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this
        is a NIH_MGC library."
        is a NIH_MGC library."

BASE COUNT  221 a 253 c 239 g 147 t 1 others
ORIGIN
Query Match 32.8%; Score 556.4; DB 14; Length 861;
Best Local Similarity 82.8%; Pred. No. 1.4e-125;
Matches 647; Conservative 0; Mismatches 132; Indels 2; Gaps 1;

QY 683 TGATGAGTGGAGTTAGACACAGTACCACTTATGAGTATGGAAGAATGA 742
    |||||
DB 14 TGATGAGTGAAGTTAGACACAGTACCACTTATGAGTATGGAAGAATGA 73
QY 743 AGTAGAGTACCTCAAGCAGTGCAGCTCATCAAGATTTTACAATCTCAAGGTCAC 802
    |||||
DB 74 AACTGTTGACTCCAGCAGGTCAGCTCATCAAGATTTTACAATCTCAAGGTCAC 133
QY 803 CGGGCCCCAGGGTCCAAAGAGTGACAGAGATCCAGGACCCCTGGCCCAACTGGCA 862
    |||||
DB 134 CTGGCCCCAGAGGTCCAAAGAGTGACAGAGATCTCAGGACCACTGGTCCAACTGGCA 193
QY 863 ACAAGGACAGAAAGAGAGAGAGGAGGAGCTTGACCACTGGCCCTGCGGTGAGAGAG 922
    |||||
DB 194 ACAAGGACAGAAAGAGAGAGAGAGGAGAGGCTGCTCCAGCTGGCCCTGCGGTGAGAGAG 253
QY 923 GCCCAATTGACAGCAGTGCCTCCCGGAGAGAGGTCGCGGCAAGAGTCTAAGGCTCC 982
    |||||
DB 254 GCACAAATTGACAGCAGTGCCTCCCGGAGAGAGGTCGCGGCAAGAGTCTAAGGCTCC 313
QY 983 AGGGCCCCAAAGGCTCCGTTGTTCCCTTGGAAGCCCGGCTCAGGGCCCAAGTGGG 1042
    |||||
DB 314 AGGGTCCCAAAAGATCTCGTGGGTCCCGCAGGAAGCCTGGCCCTCAAGACCTAGTGGG 373
QY 1043 ACCCAGAGCCCCCGGGCCCAACAGCAAGAGGAGTCCCGGCTCAGGGCCCTCTCTG 1102
    |||||
DB 374 ACCCAGAGCCACCAAGGCTCCAGCAAGAGTGGAGTCCCTGGCCCTCAGGGCCCTCTCTG 433
QY 1103 GCTTCAGGACCTTCAGAGGACCGTGGGAGAGCTGCGGTGAGACCTCGGGGAGCTG 1162
    |||||

```

```

DB 434 GCTTCCAGGGACTACAGGGCACTGTGGGTAGCCCTGGAGTACCTCGAGCTCGGGGGTTGC 493
QY 1163 CAGGCTTGCCCTGGGGTACAGGATGCGCAGGCCCCCAAGGGCCCCCGGCTCTCTGGCC 1222
    |||||
DB 494 CAGGCTTGCCAGGGGGTCCAGGATGCTGCGCTTAAGGACCACTGGCTCTCCAGGCC 553
QY 1223 CATCAGAGGAGGTGGTCCCTGGCCCTGCGAGTAAAGCCAGCCCGGAGAGACA 1282
    |||||
DB 554 CCTCAGAGAGCAATGAGAGCCATTGGCTCTGCAAAATGACCAACCCCAACATCAGAGTCA 613
QY 1283 ATGGCTCCCGGCTCTCAGTGAAGACTTCACACAAATCTCTATTTTCAGTTGAGA 1342
    |||||
DB 614 ACGAGTCTCCGCTCAGTGAAGAACTTCACAAATTAATGCTATTTTCTTGGCA 673
QY 1343 AGCAAAATTTTGAAGATGCAAGC--TTTCTGTGAAGCAACTCTTCACATCTGTTT 1400
    |||||
DB 674 AAAAAATTTTGAAGATGCTACCTTTTCTGGAACAAATCTCCCATCGGGTTT 733
QY 1401 CATTAACACTAGAGAGAGACAGCAATGATAAAAAAGATGATGAGAGAGAGCA 1460
    |||||
DB 734 CATTAACCCCAAGAAAGAAAGAGAGATGATAAAAACCTTACCTGGGGAGCCAGAAAC 793
QY 1461 C 1461
DB 794 C 794

RESULT 7
LOCUS      B0674807
DEFINITION B0674807 668 bp mRNA linear EST 15-JUL-2002
ACCESSION AGENCOURT_8035172 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6212672
VERSION    B0674807
KEYWORDS   B0674807.1 GI:21785641
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL    NIH-MGC http://mgi.nci.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT     Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLM2377 row: c column: 09
            High quality sequence stop: 656.
FEATURES
    source
        1..668
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:6212672"
        /clone_1ib="NIH_MGC_102"
        /tissue_type="epidermoid carcinoma, cell line"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
        Site_2: EcoRI; cDNA made by oligo-dT priming.
        Directionally cloned into EcoRI/XhoI sites using the
        following 5' adaptor: GGCACGAG(C). Library constructed
        by Ling Hong in the laboratory of Gerald M. Rubin
        (University of California, Berkeley) using ZAP-cDNA
        synthesis kit (Stratagene) and Superscript II RT (Life
        Technologies). Note: this is a NIH_MGC library."

BASE COUNT  216 a 148 c 167 g 137 t
ORIGIN

```


Query Match 31.0%; Score 525.4; DB 14; Length 668;
 Best Local Similarity 99.8%; Pred. No. 4.9e-118;
 Matches 526; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1169 TCCCTGGGGTACCAGGACGCGCCCAAGGCCCCCGGCTCTCTGCGCCCTACG 1228
 DB 1 TCCCTGGGGTACCAGGACGCGCCCAAGGCCCCCGGCTCTCTGCGCCCTACG 60

OY 1229 GAGCGGTGGTGGCCCTGGCCCTGCAAAATGAGCCAAACCCGCGACCGAGGACAAATGGCT 1288
 DB 61 GAGCGGTGGTGGCCCTGGCCCTGCAAAATGAGCCAAACCCGCGACCGAGGACAAATGGCT 120

OY 1289 GCGCCGCTCCTCAGTGAAGAACTTACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAA 1348
 DB 121 GCGCCGCTCCTCAGTGAAGAACTTACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAA 180

OY 1349 TTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAGCTTTCACATCTTTGTTTCATTAACA 1408
 DB 181 TTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAGCTTTCACATCTTTGTTTCATTAACA 240

OY 1409 CTAGAGAGAACAGCAATGATTAATAAACAATGATGTTAGGAGAGAGACCCACTGGATCG 1468
 DB 241 CTAGAGAGAACAGCAATGATTAATAAACAATGATGTTAGGAGAGAGACCCACTGGATCG 300

OY 1469 GCCTCAGACACTCAGAGCGTGAATAATGATGAGTGGCTGATGGACATCTCCAGACT 1528
 DB 301 GCCTCAGACACTCAGAGCGTGAATAATGATGAGTGGCTGATGGACATCTCCAGACT 360

OY 1529 ACAAATTTGGAAGCTGAGACAGCCGATTAAGTGGGCTCATGGCCATGGCCAGAGAGAG 1588
 DB 361 ACAAATTTGGAAGCTGAGACAGCCGATTAAGTGGGCTCATGGCCATGGCCAGAGAGAG 420

OY 1589 ACTGAGTGGGTTGATTAATGCTGGGCACTGGAAGCATTTTCAATCTGAAGACGTCAATA 1648
 DB 421 ACTGAGTGGGTTGATTAATGCTGGGCACTGGAAGCATTTTCAATCTGAAGACGTCAATA 480

OY 1649 ACTTCATTTTGGCAAAAGACAGAGAGACAGTACTGTCATCTGCATTA 1695
 DB 481 ACTTCATTTTGGCAAAAGACAGAGAGACAGTACTGTCATCTGCATTA 527

RESULT 8
 BE805750/c 511 bp mRNA linear EST 12-JAN-2001
 LOCUS OVI-C10173-071100-464-c02 C10173 Homo sapiens CDNA, mRNA sequence.
 DEFINITION BE805750
 ACCESSION BE805750.1 GI:12134739
 VERSION
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 511)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.J.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SF,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=OVI16t2=OVI-C10173-071100-464-c02&t3=2000-11-07&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 492.
 Location/Qualifiers
 1..511
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="C10173"
 /dev_stage="Adult"
 /note="Organ: colon; lins: Vector: puc18; Site:1: SmaI;
 Site:2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 94 a 113 c 139 g 165 t
 ORIGIN

Query Match 28.8%; Score 488.6; DB 12; Length 511;
 Best Local Similarity 98.4%; Pred. No. 4.7e-109;
 Matches 505; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

OY 27 CGTGCTCAGGCAATCTCAGAACCAATGATTTCTCAATATGTCATCATGAACCT 86
 DB 511 CGTGCTCAGGCAATCTCAGAACCAATGATTTCTCAATATGTCATCATGAACCT 452

OY 87 CAACAACCTGAACCTGACCCAGGTGAGAGAGAAAGCTCATCAGCAATCTGACGGCT 146
 DB 451 CAACAACCTGAACCTGACCCAGGTGAGAGAGAAAGCTCATCAGCAATCTGACGGCT 392

OY 147 TGTGATGACACAAGCGAGGCTATCCAGGATTCAGAAAGCACTTCAAAATCTGACGA 206
 DB 391 TGTGATGACACAAGCGAGGCTATCCAGCAATCAAGAAAGCACTTCAAAATCTGACGA 332

OY 207 GGTATTTTCTTCAAGCCAAAGAGACAGGATTTGCTGAAGAGAAAGTGACAGCTTGA 266
 DB 331 GGTATTTTCTTCAAGCCAAAGAGACAGGATTTGCTGAAGAGAAAGTGACAGCTTGA 272

OY 267 GAGCGTGGCTGCCAACACACTGCGTTGGCCAAAGCCAAACAGACACCCCTGGAGATAT 326
 DB 271 GAGCGTGGCTGCCAACACACTGCGTTGGCCAAAGCCAAACAGACACCCCTGGAGATAT 212

OY 327 GAACAGCCAGCTCACTCATTTACAGAGTCAGATGGAGAACATCACCATATCTCTCAAGC 386
 DB 211 GAACAGCCAGCTCACTCATTTACAGAGTCAGATGGAGAACATCACCATATCTCTCAAGC 152

OY 387 CAACGAGCAAGCACTGAAGACCTGAGAGCTTACCAAGAATGACAGAAATGAGACAGC 446
 DB 151 CAACGAGCAAGCACTGAAGACCTGAGAGCTTACCAAGAATGAGACAGAAATGAGACAGC 92

OY 447 CATCAAGTTCAACCAACTGGAGAACGCTTCCAGCTCTTTGAGACGAGATATTTGAACT 506
 DB 91 CATCAAGTTCAACCAACTGGAGAACGCTTCCAGCTCTTTGAGACGAGATATTTGAACT 32

OY 507 CATTCAGATATTCAGTTACACAGCCACCAACCT 539
 DB 31 CATTCAGATATTCAGTTACACAGCCACCAACCT 3

RESULT 9
 B0127513 581 bp mRNA linear EST 19-APR-2002
 LOCUS B0127513
 DEFINITION i160h06.y1 Melton Normalized Mixed Mouse Pancreas 1 N1-MWS1 Mus
 musculus cDNA clone IMAGE:5946107 5' similar to TR:090M15 090M15
 CSR2.1; mRNA sequence.
 ACCESSION B0127513
 VERSION B0127513.1 GI:20201424
 KEYWORDS
 SOURCE EST.
 ORGANISM house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 581)
 AUTHORS Melton, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kaestner, R., Lemishka, I., Seacore, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistein, A., Schmitt, A., Theising, B., Riter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T., Jackson, Y., and Bowers, Y.
 TITLE Endocrine Pancreas Consortium
 JOURNAL Unpublished (2000)
 COMMENT Other ESTs: 1160h06.x1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohpc.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center This clone is available royalty-free through LBNL; please contact the IMAGE consortium (info@image.llnl.gov) for further information
 Seq primer: -40RP from Gibco
 High quality sequence stop: 446.

FEATURES
 source
 Location/Qualifiers
 1..581
 /organism="Mus musculus"
 /strain="ICR"
 /db_xref="taxon:10090"
 /clone_1lb="Melton Normalized Mixed Mouse Pancreas 1 N1-MMS1"
 /sex="Both for embryonic & newborn, male for adult and adult islet"
 /dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn, adult, mixed"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; Five libraries representing E10.5/12.5 pancreatic bud, E16.5 pancreas, newborn pancreas, adult pancreas, and adult islets of langerhans were separately constructed using Superscript Plasmid Library kit (Life Technologies). cDNA was made by oligo-dT priming and size-selected by column fractionation. Libraries were amplified once on solid support and plasmid DNA from each library was prepared and mixed in equal amounts. The mixed library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded mixed library plasmid DNA was mixed with 5 micrograms PCR product representing mixed library inserts and hybridized to an EcoT of 6. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 189 a 159 c 114 g 119 t
 ORIGIN

Query Match 28.3%; Score 480.2; DB 14; Length 581;
 Best Local Similarity 89.2%; Pred. No. 5.7e-107;
 Matches 518; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 86 TCACACACCTGAACTGACCCAGTGCAGAGAGAACTCATACGAATCTGCAGCGGT 145
 |||||
 Db 1 TCACACACCTGTAACCTGACCCAGTGCAGAGAGAACTCATACGAATCTGCAGAGT 60
 QY 146 CTGTGATGACACAGCAGGCTATCGAGCAATCAAGAACGACTTTCAAAATCTGCAGC 205
 |||||
 Db 61 CTGTGATGACACAGCAGGCTATCGAGCAATCAAGAACGACTTTCAAAATCTGCAGC 120
 QY 206 AGGTTTTTCTTCAAGCCAAAGACAGGATGGCTGGAAGGAAGGACGAGAGCTTGC 265
 |||||
 Db 121 AGGTTTTTCTTCAAGCCAAAGACAGGATGGCTGGAAGGAAGGACGAGAGCTTGC 180

QY 266 AGACGCTGCTGCCAACACTGTGCGTTGGCCAAAGCCAAACAGACACCTTGAGAGATA 325
 |||||
 Db 181 AGACATTCGCTGCCAACACTGTGCGTTGGCCAAAGCCAAACAGACACCTTGAGAGATA 240
 QY 326 TGAACGCCAGCTCACTCATCTTTCACAGAGTGAAGAACATCACCATCTCTCAAG 385
 |||||
 Db 241 TGAATACCGCTCAGCTCATCTTTCACAGAGTGAAGAACATCACCATCTCTCAAG 300
 QY 386 CCAACGACGAAACCTGAAAGACCTGACGAGCTTACACAAAGATGACAGATAAGACAG 445
 |||||
 Db 301 CCAACGACGAGACGCTGAAGACCTTACAGACTTACACAAAGATGACAGATAAGACAG 360
 QY 446 CCAATCAAGTTCAACCAACTGAGAGACGCTTCCAGCTCTTTGAGAGGATATTGTGACA 505
 |||||
 Db 361 CTGTCAAGTTCCAGCCCAACTTGAAGAACGCTTCCAGGCTTTTGAAGACAGATATTGTGACA 420
 QY 506 TCATTGCAATATCAGTTATACAGCCACACCGGCGGAGCGTCGACAGCAATCTTAATG 565
 |||||
 Db 421 TCATTGCAATATCAGTTATACAGCCACACCGGCGGAGCGTCGACAGCAATCTTAATG 480
 QY 566 AAGTCAGACCACTTGCACAGATACCTTTACCAACACACAGATGATCTGACCTCTTGA 625
 |||||
 Db 481 ATGTTAGACCATGATGACAGACACCTTGAACAGACAGACGATGACCTGACCTCTTGA 540
 QY 626 ATAAATACCTGCGCCACATCCGTTTGATTCGTTCTCTC 666
 |||||
 Db 541 ATAAACACACTGATCAACATCCGCTTGATTCCTCTCTC 581

RESULT 10
 AL568743/c AL568743 967 bp mRNA 1linear EST 16-FEB-2001
 LOCUS
 DEFINITION AL568743 L1L_FL002_PL1 Homo sapiens cDNA clone CSDBE005YH04 3 prime
 , mRNA sequence.
 ACCESSION AL568743
 VERSION AL568743.1 GI:12923387
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 967)
 AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source
 Location/Qualifiers
 1..967
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CSDBE005YH04"
 /clone_1lb="L1L_FL002_PL1"
 /lab_host="DH10B"
 /note="Organ: Placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(47) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 222 a 221 c 217 g 290 t 17 others
 ORIGIN

Query Match 28.0%; Score 475.4; DB 9; Length 967;
 Best Local Similarity 96.7%; Pred. No. 1.1e-105;
 Matches 504; Conservative 3; Mismatches 9; Indels 5; Gaps 2;

9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 3528
 Fax: 301 838 0208
 Email: johng@ligr.org
 Plate: 115

Seq primer: Reverse.

FEATURES
 Location/Qualifiers

source

1..552
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE_respositories"
 /note="Vector: pBluescriptSKm"

BASE COUNT 176 a 112 c 143 g 121 t
 ORIGIN

Query Match 25.8%; Score 437.2; DB 10; Length 552;

Best Local Similarity 99.1%; Pred. No. 2e-96;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1243 CTGGCCCTGCAGATGAGCCACCCGCGACCGAGAGCAATGGCTGCCCTCCTCCTG 1302
 DB 1 CTGGCCCTGCAGATGAGCCACCCGCGACCGAGAGCAATGGCTGCCCTCCTCCTG 60

QY 1303 AAGAACTTCACAGCAAAATCTCTACTATTTTCAGTTGAGAGAAATTTTGAAGATGCA 1362
 DB 61 AAGAACTTCACAGCAAAATCTCTACTATTTTCAGTTGAGAGAAATTTTGAAGATGCA 120

QY 1363 AAGCTTTTGTGTAGAGCAAGTCTTCATCTTTTTCATTAACACTAGAGGAAACAG 1422
 DB 121 AAGCTTTTGTGTAGAGCAAGTCTTCATCTTTTTCATTAACACTAGAGGAAACAG 180

QY 1423 CAATGATATAAAACAGATGTAGAGAGAGAGACCACTGATCGGCTCAGACACTCA 1482
 DB 181 CAATGATATAAAACAGATGTAGAGAGAGAGACCACTGATCGGCTCAGACACTCA 240

QY 1483 GACCGTGAATAATGATGAGATGGCTGATGGACATCTCCACATCAAAATTTGAAA 1542
 DB 241 GACCGTGAATAATGATGAGATGGCTGATGGACATCTCCACATCAAAATTTGAAA 300

QY 1543 GCTGGACAGCCGATTAAGTGGGTTCATGCGCATGGGCGACAGAGACACTGCTGGGTTG 1602
 DB 301 GCTGGACAGCCGATTAAGTGGGTTCATGCGCATGGGCGACAGAGACACTGCTGGGTTG 360

QY 1603 ATTATGCTGGGCACTGAGAAAGATTTCCAAATGTGAAGACGTCATAAATAC-TTCATTTGCGA 1661
 DB 361 ATTATGCTGGGCACTGAGAAAGATTTCCAAATGTGAAGACGTCATAAATTTGCGA 420

QY 1662 AAAAGACAGGAGACAGTACTGATCTGATTA 1695
 DB 421 AAAAGACAGGAGACAGTACTGATCTGATTA 454

RESULT 13
 BE290299 601 bp mRNA linear EST 26-OCT-2000
 LOCUS 601089246F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3484163 5',
 DEFINITION mRNA sequence.

ACCESSION BE290299
 VERSION BE290299
 KEYWORDS BE290299.1 GI:9171250
 EST

SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 601)
 TITLE NIH-MGC http://mgi.nci.nih.gov/
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: L1AM8516 row: k column: 12
 High quality sequence stop: 551.

FEATURES
 Location/Qualifiers

source

1..601
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:3484163"
 /clone_1ib="NCI_CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site.1: SalI;
 Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

BASE COUNT 156 a 161 c 179 g 105 t
 ORIGIN

Query Match 24.5%; Score 415.8; DB 10; Length 601;

Best Local Similarity 84.8%; Pred. No. 3.7e-91;

Matches 501; Conservative 0; Mismatches 87; Indels 3; Gaps 3;

QY 865 AAGGACAGAAAGAGAGAGAGGAGGAGCCCTGACACCTGGCTCGGTGAGAGAGC 924
 DB 9 AAGTGAACAGAAAGAGAGAGGAGAGCCCTGACACCTGGCTCGGTGAGAGAGC 68

QY 925 CCAATTGACACAGCTGATCCCTCCGAGAGAGCTGGGCGCAAGATCTAAGGCTCCAG 984
 DB 69 ACAATTGACACAGCTGATCCCTCCGAGAGAGCTGGGCGCAAGATCTAAGGCTCCAG 128

QY 985 GGCCTCAAGAGCTCCCTGCTCCCTGGG-AAGCCGCGCTCAGAGCCCTCAGTGGGA 1043
 DB 129 GCTCCAAAGAGCTCCCTGCTCCCTGGGTAAGCTGCGCTCAAGAGACTGATGGGA 188

QY 1044 CCGAGGCCCCCGGGCCACACAGCAAGAGAGACCTCCGCGCTCAGAGGCTCTCGG 1103
 DB 189 CCGAGGACACAGCTCCACAGCAAGAGAGACTCCCTGCGCTCAGAGGCTCTCGG 248

QY 1104 CTTCAGAGGACTTCAGAGGACCCCTGGGAGACCTGGGCTCTGAGACTCGGGAGTGGC 1163
 DB 249 CTTCAGAGGACTTCAGAGGACCTGGGAGACCTGGGAGTGGGAGTGGGAGTGGC 308

QY 1164 AGGCTTGCTGGGGTACAGGACATGCCAGGCCCAAGAGGCCCTCGGCTCTGGGCC 1223
 DB 309 AGGCTTGCTG-AGGCTGCAAGGACATGCCCTGGGCTTAAGGAGCCACTGGGCTCTGAGGCC 367

QY 1224 ATCAGGAGGCGGTGGTCCCTGCGCTGCGAGATGAGCCACCCGCGACCGGAGAGCA 1283
 DB 368 CTGAGGAGCAATGAGAGCAATGCTCTGCGAGATGAGCCACCCGCGACCGGAGAGCA 427

QY 1284 TGCGTCCCGGCTCCTGAGAGACTTCACAGCAAAATGCTACTATTTTTCAGTTGAG 1343
 DB 428 CGGATGCGCGCTCCTGAGAGAACTTCACAGTAATATGCTACTATTTTTCAGTTGAG 487

QY 1344 AGAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGCAAGCTCTTCACATCTTTTCAT 1403
 DB 488 AGAAATTTTGAAGATGCAAAAGCTTTTCTGTGAAGCAAAATCTCTTCATCTTTTCAT 546

QY 1404 AAACACTAGAGAGAGAGCAAGATGATTAATAAAGATGATGAGAGAGA 1454
 DB 547 AAACCTAAGAGAGAGAGCAAGATGATGATTAATAAAGATGATGAGAGAGA 597

RESULT 14
 BF856923/c 410 bp mRNA linear EST 16-JAN-2001
 LOCUS BF856923
 DEFINITION QV1-FT0201-021100-453-e01 FT0201 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF856923 GI:12244667
 VERSION BF856923.1
 KEYWORDS EST
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 410)
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=0V1&c2=0V1-FT0201-021100-453-e01&t3=2000-11-02&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 410.
 Location/Qualifiers
 1..410
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="FT0201"
 /dev_stage="Adult"
 /note="Organ: prostate tumor; Vector: puc18; Site:1; SmaI; Site:2; SmaI; A mini-library was made by cloning products derived from ORESTES PCR (O.S. Letters Patent Application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 73 a 93 c 109 g 135 t
 ORIGIN
 Query Match 23.4%; Score 395.8; DB 12; Length 410;
 Best Local Similarity 99.3%; Pred. No. 2.5e-86;
 Matches 408; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 110 GCAGAGATAGAACAGCCATCAAGTTCACCAACTGGAGGACGCTTCAGCTCTTGAG 51
 Oy 490 ACGGATATTGTGTAACATCATTTAGCAATATTCAGTTCACAGCCACCACCTG 540
 Db 50 ACGGATATTGTGTAACATCATTTAGC-ATACAGATTACAGCCACCACCTG 1
 RESULT 15
 LOCUS BB248064 638 bp mRNA linear EST 23-OCT-2001
 DEFINITION BB248064 RIKEN full-length enriched, 7 days neonate cerebellum Mus musculus cDNA clone A730023E20 3', mRNA sequence.
 ACCESSION BB248064
 VERSION BB248064.2 GI:16355610
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 638)
 AUTHORS Arikawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyata, T., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Arikawa, T., et al. 2001)
 Unpublished (2001)
 On Jul 6, 2000 this sequence version replaced g1:8940810.
 CONTACT Yoshinori Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gs.riken.go.jp
 URL: <http://genome.gsc.riken.go.jp/>
 Carninci, P., Shibata, Y., Hayashizaki, Y., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Location/Qualifiers
 1..638
 /organism="Mus musculus"
 /db_xref="taxon:10090"

```

BASE COUNT      212 a      134 c      146 g      146 t.
ORIGIN
/clone="A730023E20"
/clone_lib="RIKEN full-length enriched, 7 days neonate
cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/notes="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer 15'
GAGGAGAGAGAGATCCGAGGCTCTTTTCTTTTCTTTTCTTTVN 3', cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence 15' GAGGAGAGAGATCTCGAGTTAAATTAATTCACCCCCCCCCC
3'. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pbluescript KS(+) after bulk excision from Lambda
FLC I."
```

| | | | | |
|-----------------------|----------------|-------------------|----------|------------|
| Query Match | 21.88 | Score 369.4 | DB 10 | Length 638 |
| Best Local Similarity | 84.58 | Pred. No. 9.4e-80 | | |
| Matches 415 | Conservative 0 | Mismatches 76 | Indels 0 | Gaps 0 |

| | | | |
|----|------|--|------|
| Oy | 1205 | CCCCGGGCCCCCTGGGCCCCATTCAGAGAGAGGGGTGGTGGCCCTGGGCCCCGCGAATGAGCCAA | 126 |
| Db | 1 | CACCTGGGCCCTCCAGAGGCCCTCCAGAGACATGAGGCCATTGGCTCTGCAGAAATGAAACCA | 60 |
| Oy | 1285 | CCCCGGACCGGAGACATGCGTGCCTCCCTCATTGAGAAACTTTCACGACCAAAATGCT | 1324 |
| Db | 61 | CCCCGAGCATGAGAGGTCCACGGATGTCCGCTCCACGGAAGAACTTCACAGATTAATGCT | 120 |
| Oy | 1325 | ACATATTTTTCAGTGGAGAAAGAAATTTTGGAGATGCCAAAGCTTTCTGTGSAAGCAAGT | 1384 |
| Db | 121 | ACATATTTTTCATTGSAAGAAAAATTTTGGAGATCTTAAGCTTTTCTGTGAAAGCAAAAT | 180 |
| Oy | 1385 | CTTCACATCTTGTGTTTTCATTAACACTAGAGAGGACAGCAATGGATTAAGAAAAACAGATGG | 1444 |
| Db | 181 | CTTCCCATCTCGTTTTCATTAACTCAAGAGAGAAAGAACAGCAATGGATTAAGAAAAAGCATACCG | 240 |
| Oy | 1445 | TAGGAGAGAGAGCCACTGGATGCGGCTCACAAGACTCAGAACGCTGAAGATGAATGGAAAT | 1504 |
| Db | 241 | TGGGGAAGAGAAAGCCATTGGATGCGGCTCACAAGACTCAAGAACAGGAAACCGAATGGAAGT | 300 |
| Oy | 1505 | GGCTGGATGGACATCTCCAGACTACAAAAAATTTGGAAGACTGGACGCCGGAATTAATGGG | 1564 |
| Db | 301 | GGCTAGACGGGCTACCTGTTGATTACAAAAACTGGAAGACTGGACAAACCAAGATTAATGGG | 360 |
| Oy | 1565 | GTCATGGCCATGGGGCCAGAGAAAGACTGTGCTGGGTGATTATGCTGGGCGAGTGAACG | 1624 |
| Db | 361 | GCAGTGGCCATGGGGCCAGAGAAAGACTGTGCTGGGTGATTATACGAGGACAGTGGAAATG | 420 |
| Oy | 1625 | ATTTTCCATGTGAAGAGCTGAATTAACCTCATTTTGCAGAAAAAGACAGAGAGCAAGTACTGT | 1684 |
| Db | 421 | ACTTTCAGTGTGAAGAAATCAATTAAGCTTATTTGTGAGAAAGAAAGGAGGAGCAATCAAT | 480 |
| Oy | 1685 | CATCTGCATTA 1695 | |
| Db | 481 | CATCCATATTTA 491 | |

Search completed: March 21, 2003, 08:07:57
Job time : 2240.7 secs

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Run on: March 21, 2003, 01:02:25 ; Search time 3471.92 Seconds

Title: US-09-763-712A-1_COPY_325_1695
 Portfast Score: 1271

Scoring table: IDENTITY_NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

| | | | |
|---------|--------|---------|---|
| Maximum | DB seq | length: | 0 |
| Minimum | DB seq | length: | 0 |

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

| | | |
|-----|---------------|---|
| 1: | gd_da: | * |
| 2: | gd_htg: | * |
| 3: | gb_in: | * |
| 4: | gb_cm: | * |
| 5: | gb_ov: | * |
| 6: | gb_pat: | * |
| 7: | gb_ph: | * |
| 8: | gb_pl: | * |
| 9: | gb_pr: | * |
| 10: | gb_ro: | * |
| 11: | gb_stg: | * |
| 12: | gb_sy: | * |
| 13: | gb_un: | * |
| 14: | gb_vl: | * |
| 15: | em_ba: | * |
| 16: | em_fun: | * |
| 17: | em_hum: | * |
| 18: | em_in: | * |
| 19: | em_mu: | * |
| 20: | em_om: | * |
| 21: | em_or: | * |
| 22: | em_ov: | * |
| 23: | em_pat: | * |
| 24: | em_ph: | * |
| 25: | em_pl: | * |
| 26: | em_ro: | * |
| 27: | em_sts: | * |
| 28: | em_un: | * |
| 29: | em_vl: | * |
| 30: | em_htg_hum: | * |
| 31: | em_htg_inv: | * |
| 32: | em_htg_other: | * |
| 33: | em_htg_mus: | * |
| 34: | em_htg_pin: | * |
| 35: | em_htg_mem: | * |
| 36: | em_htg_mam: | * |
| 37: | em_htg_vrt: | * |
| 38: | em_sy: | * |
| 39: | em_hgo_hum: | * |
| 40: | em_hgo_mus: | * |
| 41: | em_hgo_other: | * |

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

| Result | Query | Description |
|--------|--------------------------|-------------|
| No. | Score Match Length DB ID | |
| ----- | | |

| | | | | | |
|------|--------|-------|--------|----|-------------|
| 1 | 1377 | 100.0 | 2983 | 5 | ABO051145 |
| 2 | 1369.4 | 99.9 | 2641 | 6 | ABX047353 |
| 3 | 1369.4 | 99.9 | 3058 | 6 | ABO385118 |
| 4 | 1357.4 | 99.0 | 2005 | 6 | AXA554442 |
| 5 | 1357.4 | 99.0 | 2005 | 6 | AXA490920 |
| 6 | 1251.4 | 91.3 | 1886 | 6 | HSM802985 |
| 7 | 1086.2 | 79.2 | 2637 | 10 | ABO78434 |
| 8 | 1083 | 79.0 | 3291 | 10 | ABO585119 |
| 9 | 960.4 | 70.1 | 4330 | 10 | ABO521103 |
| C 10 | 488.8 | 35.7 | 169088 | 2 | AC016128 |
| C 11 | 488.8 | 35.7 | 188439 | 2 | AP000915 |
| 12 | 484.6 | 35.3 | 71044 | 2 | AC024368 |
| 13 | 474 | 34.6 | 178022 | 2 | AP000900 |
| 14 | 474 | 34.6 | 187635 | 2 | AP001022 |
| 15 | 462 | 33.7 | 188255 | 2 | AP000939 |
| C 16 | 437.4 | 31.9 | 188255 | 2 | AP000939 |
| C 17 | 389.2 | 28.4 | 182029 | 2 | AC114677 |
| C 18 | 389.2 | 28.4 | 193208 | 2 | AC102618 |
| C 19 | 388 | 28.3 | 182029 | 2 | AC114677 |
| 20 | 388 | 28.3 | 193208 | 2 | AC102618 |
| 21 | 378.4 | 27.6 | 130763 | 2 | AC112416 |
| 22 | 235.6 | 17.2 | 3636 | 6 | ABO07829 |
| 23 | 235.6 | 17.2 | 3685 | 6 | E32511 |
| 24 | 235.6 | 17.2 | 3810 | 6 | E33509 |
| 25 | 222.8 | 16.3 | 2215 | 10 | BC026446 |
| C 26 | 166.8 | 10.7 | 178022 | 2 | AP000900 |
| C 27 | 146.8 | 10.7 | 187635 | 2 | AP001022 |
| C 28 | 129 | 9.4 | 6156 | 10 | ABO09993 |
| 29 | 128.8 | 9.4 | 152181 | 2 | AC126444 |
| 30 | 128.8 | 9.4 | 219809 | 2 | AC125058 |
| 31 | 128.2 | 9.4 | 2193 | 3 | PALC01A |
| 32 | 125.8 | 9.2 | 6076 | 10 | AF272661 |
| 33 | 124.2 | 9.1 | 1041 | 4 | AF138883 |
| 34 | 124.2 | 9.1 | 2053 | 5 | ABO08374 |
| 35 | 124.2 | 9.1 | 4502 | 5 | ABO52836 |
| 36 | 123 | 9.0 | 1711 | 6 | ABO07830 |
| 37 | 123 | 9.0 | 1877 | 6 | E32510 |
| 38 | 122.6 | 8.9 | 4995 | 10 | MM016789 |
| C 39 | 122 | 8.9 | 95002 | 2 | AF233390 |
| C 40 | 122 | 8.9 | 134514 | 2 | AC025192 |
| C 41 | 122 | 8.9 | 157382 | 2 | AC013643 |
| C 42 | 122 | 8.9 | 175382 | 2 | AF235104 |
| C 43 | 121.6 | 8.9 | 810 | 6 | AB014116 |
| C 44 | 121.4 | 8.9 | 4628 | 4 | ABO08683 |
| 45 | 121.4 | 8.9 | 5676 | 6 | E07265 |
| | | | | | cnDNA encod |

ALIGNMENTS

| | |
|------------|---|
| RESULT 1 | |
| AB005145 | |
| LOCUS | AB005145 2983 bp mRNA linear PRI 21-NOV-2001 |
| DEFINITION | Homo sapiens CL-P1 mRNA for collectin placenta 1, complete cds. |
| ACCESSION | AB005145 |
| VERSION | AB005145.1 GI:17026100 |
| KEYWORDS | . |
| SOURCE | Homo sapiens female tissue_1lb:placenta cDNA to mRNA. |
| ORGANISM | Homo sapiens |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; |
| | Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. |
| REFERENCE | 1 |
| AUTHORS | Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H., Sakai,Y., Fukuchi,A., Sakamoto,T., Itabe,H., Suzutani,T., Ogasawara,M., Yoshida,I. and Wakamiya,N. |

TITLE The membrane-type collectin CL-P1 is a scavenger receptor on
JOURNAL J. Biol. Chem. 276 (47), 44222-44228 (2001)
MEDLINE 21570232
REFERENCE 2 (bases 1 to 2983)
AUTHORS Ohtani,K.
TITLE Direct Submission
SUBMITTED (25-JUN-1997) Katsuki Ohtani, Asahikawa Medical College,
Department of Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa,
Hokkaido 078-8510, Japan (E-mail: ohtani@asahikawa-med.ac.jp,
Tel: +81-166-68-2393, Fax: +81-166-68-2399)
FEATURES
Source
1..2983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="female"
/tissue.lib="placenta"
1..2983
/gene="CL-P1"
71..2299
/gene="CL-P1"
/codon_start=1
/product="collectin placenta 1"
/protein_id="BAB72147.1"
/db_xref="GI:17026101"
/translation="MKDFAEEEEVSGFYKRFQIGEQCTCKNNMALKFSILLY
ILCALITIVAILGYKVEKMDVVTGMEVSRQYDQDKLTAVSDIKLQDQKKA
STNSELSTERSDILDLROOLREITERKSNKDLKLSGDSLAVROSQKLTLEN
SFLITVNTKLOANGVYNTLNODTSVLQGNLONOMYSHNVVIMNLTNNLTPVOOR
LITNLOASVDNDSOALORIKNDPNOAOYFLOAKKPTDMLKEKVSLOTLANNSLA
KANNDLEBNSLNSFTGOMENITTSQANEONLKDLDLKHDAENRIRAIKNOIE
RFQLEFDIVNIIISNISTYRAHLRLTSLNENRTCTDPLTKHTDVLTLNNTLANI
RLDSVSLRMQODLMRSRLDEYANLVSIMENKLVSKHQLIKNTILOGPGRGP
RGRSGSQGPPTGPNKGOKGEKPEPPEPAGRCPIGAPPEGEKSGSKSQPK
GSRGSGKPGPQSGDPGPPGPKGELGEPGPGFQGLQGTVEPGVGPGRGLG
LPGVPGMPGKGPBPSPGPGSAVPLALNEPAPEDNCCPPEHMKNEFDKCYEVE
KEIFEDAKLFCEDKSSHIYFINTREDOQTKKQVGRRESHIGLITSERENEMKMDG
TSPDYNNMKAQPDNMNGHGHGPEDCAGLTYAGWMDPCEDEVNFIKEDRETVLSS
AL"

BASE COUNT 914 a 707 c 703 g 659 t
ORIGIN

Query Match 100.0%; Score 1371; DB 9; Length 2983;
Best Local Similarity 100.0%; Pred. No. 2.2e-308;
Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACAGCAGCTCACTCATTTCACAGCTGATGAGAAACATCACACTATCTCA 60
|||||
Db 926 ATGAACAGCAGCTCACTCATTTCACAGCTGATGAGAAACATCACACTATCTCA 985
QY 61 GCCAAGCAGCAGAACTGAAAGACCTGACAGACTTACACAAAGATGCAGAAATAGA 120
|||||
Db 986 GCCAAGCAGCAGAACTGAAAGACCTGACAGACTTACACAAAGATGCAGAAATAGA 1045
QY 121 GCATCAAGTTCAACCACTGAGAGACGCTTCCAGCTCTTTGAGAGGATATTGTAA 180
|||||
Db 1046 GCATCAAGTTCAACCACTGAGAGACGCTTCCAGCTCTTTGAGAGGATATTGTAA 1105
QY 181 ATCATAGCAATATCAGTTACACAGCCACACCTGGGAGCGTGACACAAATCTAA 240
|||||
Db 1106 ATCATAGCAATATCAGTTACACAGCCACACCTGGGAGCGTGACACAAATCTAA 1165
QY 241 GAAGTCAGACCACTTGCACAGATACCTTACCAACACACAGATGATGACTGCTCTG 300
|||||
Db 1166 GAAGTCAGACCACTTGCACAGATACCTTACCAACACACAGATGATGACTGCTCTG 1225
QY 301 AATTAATACCTGGCCAAACATCCGTTGGATTCTGTTCTCAGATGACAAAGATTGG 360
|||||
Db 1226 AATTAATACCTGGCCAAACATCCGTTGGATTCTGTTCTCAGATGACAAAGATTGG 1285
QY 361 ATGAGGTCGAGGTTAGACACTGAAGTACCACTATCATGATATGAGAAAGAAATAG 420
|||||
Db 1286 ATGAGGTCGAGGTTAGACACTGAAGTACCACTATCATGATATGAGAAAGAAATAG 1345

QY 421 CTAGTAGACTCCAAAGCATGTGCTCAGCTCATCAAGAATTTTACAATACTCAAGGTCCACG 480
|||||
Db 1346 CTAGTAGACTCCAAAGCATGTGCTCAGCTCATCAAGAATTTTACAATACTCAAGGTCCACG 1405
QY 481 GGGCCCAAGGGGTCCAGAGGTGACAGAGATCCAGAGGACCCCTGGCCCACTGGCAAC 540
|||||
Db 1406 GGGCCCAAGGGGTCCAAAGGTGACAGAGATCCAGAGGACCCCTGGCCCACTGGCAAC 1465
QY 541 AAGGACAGAAAGAGAGAGAGGGGAGACCTGAGACCACTGCGCCTGGGGGTGAGAGGCG 600
|||||
Db 1466 AAGGACAGAAAGAGAGAGAGGGGAGACCTGAGACCACTGCGCCTGGGGGTGAGAGGCG 1525
QY 601 CCAATTGGACACAGTGTGCTCCCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCAG 660
|||||
Db 1526 CCAATTGGACACAGTGTGCTCCCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCAG 1585
QY 661 GGGCCCAAGGGTCCCGTGTGCTCCCGTGGGAGACCCGGCCCTCAAGGCCCCCACTGGGGAG 720
|||||
Db 1586 GGGCCCAAGGGTCCCGTGTGCTCCCGTGGGAGACCCGGCCCTCAAGGCCCCCACTGGGGAG 1645
QY 721 CCAAGGCCCCCGGAGCCACAGGCAAGAGAGGACTCCCGGCTCAGGGCCCTCCCTGGC 780
|||||
Db 1646 CCAAGGCCCCCGGAGCCACAGGCAAGAGAGGACTCCCGGCTCAGGGCCCTCCCTGGC 1705
QY 781 TTCAGGAGACTTCAGGGCACCGTTGGGGAGCGTGGGTGCTTGACCTCGGGAGCTGCCA 840
|||||
Db 1706 TTCAGGAGACTTCAGGGCACCGTTGGGGAGCGTGGGTGCTTGACCTCGGGAGCTGCCA 1765
QY 841 GCGTGGCTGGGGTACAGGACATGCCAGGGCCCCAGGGCCCCCGGCTCCTGTGGGCCA 900
|||||
Db 1766 GCGTGGCTGGGGTACAGGACATGCCAGGGCCCCAGGGCCCCCGGCTCCTGTGGGCCA 1825
QY 901 TCAAGAGCGGTGGTGGCCCTGCGCCCTCAAAATGAGCAACCCCGGACCGGAGGACAAT 960
|||||
Db 1826 TCAAGAGCGGTGGTGGCCCTGCGCCCTCAAAATGAGCAACCCCGGACCGGAGGACAAT 1885
QY 961 GCGTGGCCGCTCAGTGGAAACCTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1020
|||||
Db 1886 GCGTGGCCGCTCAGTGGAAACCTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1945
QY 1021 GAAATTTTGAAGATGGAAGCTTTTCTGGAAGACAGCTTCACATCTGTTTTCATA 1080
|||||
Db 1946 GAAATTTTGAAGATGGAAGCTTTTCTGGAAGACAGCTTCACATCTGTTTTCATA 2005
QY 1081 AACACTAGAGAGAGACAGCAATGGATTAATAAACAAGATGCTAGGAGAGAGAGACCACTGG 1140
|||||
Db 2006 AACACTAGAGAGAGACAGCAATGGATTAATAAACAAGATGCTAGGAGAGAGAGACCACTGG 2065
QY 1141 ATGGGCTCAGACACTGAGAGCGTGAATGAATGAATGAGTGGCTGGATGGACATCTCCA 1200
|||||
Db 2066 ATGGGCTCAGACACTGAGAGCGTGAATGAATGAATGAGTGGCTGGATGGACATCTCCA 2125
QY 1201 GACTACAAAAATTGGAAGGTGACAGCGGATTAACGTGGGTATGGCCATGGGCCACAGA 1260
|||||
Db 2126 GACTACAAAAATTGGAAGGTGACAGCGGATTAACGTGGGTATGGCCATGGGCCACAGA 2185
QY 1261 GAAGACTGTGCTGGTGTGATTATGCTGGGACGTGGAACATTTCCAAATGTAAGACGTC 1320
|||||
Db 2186 GAAGACTGTGCTGGTGTGATTATGCTGGGACGTGGAACATTTCCAAATGTAAGACGTC 2245
QY 1321 AATTAATCTATTTGCGAAAAAGACAGGAGACAGTACTGTCAATCTGCATTA 1371
|||||
Db 2246 AATTAATCTATTTGCGAAAAAGACAGGAGACAGTACTGTCAATCTGCATTA 2296

RESULT 2
AX047353
LOCUS AX047353 2641 bp DNA linear PAT 15-DEC-2000
DEFINITION Sequence 39 from Patent WO0068380.
ACCESSION AX047353
VERSION AX047353.1 GI:11876599
KEYWORDS

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2641)
 AUTHORS Bandman, O., Hillman, J.L., Tang, Y.T., Lai, P., Yue, H., Baughn, M.R.,
 Lu, D.A. and Azimzal, Y.
 TITLE Extracellular matrix and adhesion-associated proteins
 JOURNAL Patent: WO 0068380-A 39 16-NOV-2000;
 Incyte Genomics, Inc. (US)
 FEATURES
 source location/Qualifiers
 1..2641
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 4586187CBI"
 BASE COUNT 823 a 628 c 608 g 582 t
 ORIGIN
 Query Match 99.9%; Score 1369.4; DB 6; Length 2641;
 Best Local Similarity 99.9%; Pred. No. 5.2e-308;
 Matches 1370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 1 ATGACAGCCAGCTCACTCATTCACAGGTGAGAGAACATCCACTATCTCTCAA 60
 567 ATGACAGCCAGCTCACTCATTCACAGGTGAGAGAACATCCACTATCTCTCAA 626
 61 GCCACGACGACGACCTGAAAGCCTGCAGACTTACACAAAGATGACAGATAGACA 120
 627 GCCACGACGACGACCTGAAAGCCTGCAGACTTACACAAAGATGACAGATAGACA 686
 121 GCCATCAAGTTCAACCACTGAGAACGCTTCACAGCTCTTTCAGAGAGATATTGGAAC 180
 687 GCCATCAAGTTCAACCACTGAGAACGCTTCACAGCTCTTTCAGAGAGATATTGGAAC 746
 181 ATCATTAAGCAATATCATTTACACAGCCACCACTCGGAGCGCTGACCAATCTAAT 240
 747 ATCATTAAGCAATATCATTTACACAGCCACCACTCGGAGCGCTGACCAATCTAAT 806
 241 GAAGTCAGGACCACTGACAGATACCTTACCAACACAGATATGACCTCTTG 300
 807 GAAGTCAGGACCACTGACAGATACCTTACCAACACAGATATGACCTCTTG 866
 301 AATAATACCTGCGCAACATCCGTTTGATTTCTCTCTAGATGCAACAAGATTG 360
 867 AATAATACCTGCGCAACATCCGTTTGATTTCTCTCTAGATGCAACAAGATTG 926
 361 ATGAGTGCAGGTTCAGACTGACAGTACCACTTATCATGTATATGAGAAATGAG 420
 927 ATGAGTGCAGGTTCAGACTGACAGTACCACTTATCATGTATATGAGAAATGAG 986
 421 CTAGTAGACTCCAGCATGCTGACATCAAGATTTTACAAATCTACAAAGTCCACCG 480
 987 CTAGTAGACTCCAGCATGCTGACATCAAGATTTTACAAATCTACAAAGTCCACCG 1046
 481 GGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCAGAGGCCCTTGCCCAATGCGCAAC 540
 1047 GGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCAGAGGCCCTTGCCCAATGCGCAAC 1106
 541 AAGGACACAAAGAGAGAGAGGGGAGCTTGACACACTGCGCCCTCGGGTGAGAGAGC 600
 1107 AAGGACACAAAGAGAGAGAGGGGAGCTTGACACACTGCGCCCTCGGGTGAGAGAGC 1166
 601 CCAATTGGACGACGCTGCTCCCGGAGAGCGTGGGGGCAAGAGATCTAAAGCTCCAG 660
 1167 CCAATTGGACGACGCTGCTCCCGGAGAGCGTGGGGGCAAGAGATCTAAAGCTCCAG 1226
 661 GGGCCCAAGAGGTCCCGTGGTTCCCTCTGGGAAACCGCGCCCTCAGAGGCCCAATGGGGAC 720
 1227 GGGCCCAAGAGGTCCCGTGGTTCCCTCTGGGAAACCGCGCCCTCAGAGGCCCAATGGGGAC 1286
 721 CCAAGGCCCCCGGGCCACCAAGGCAAGAGAGCTCCCGGCGCTCAGAGGCCCTCTCTGCG 780
 1287 CCAAGGCCCCCGGGCCACCAAGGCAAGAGAGCTCCCGGCGCTCAGAGGCCCTCTCTGCG 1346

QY 781 TTCAGGACCTTCAGGGACCGCTTGGGAGACCTGGGGTGGCTGGACTCGGGACTGCCA 840
 1347 TTCAGGACCTTCAGGGACCGCTTGGGAGACCTGGGGTGGCTGGACTCGGGACTGCCA 1406
 QY 841 GGGTTCCTGGGGTACCGAGCATGCCAGGCCCGCCCGCCCTCTCTGGCCCA 900
 1407 GGGTTCCTGGGGTACCGAGCATGCCAGGCCCGCCCGCCCTCTCTGGCCCA 1466
 QY 901 TCAGAGCGGTGTGCTCCCTGGCCCTGCAGATATGACCAACCCCGCAGCGAGACAT 960
 1467 TCAGAGCGGTGTGCTCCCTGGCCCTGCAGATATGACCAACCCCGCAGCGAGACAT 1526
 QY 961 GGTGGCCCGCTCAGTGAAGAACCTCACAGACAAATGCTATTTTCAGTTGAGAAA 1020
 1527 AGCTGCCCGCTCAGTGAAGAACCTCACAGACAAATGCTATTTTCAGTTGAGAAA 1586
 QY 1021 GAAATTTTTCAGAGTCAAAAGCTTTCTGTGAAGACAGTCTTCACATCTTGTTCATA 1080
 1587 GAAATTTTTCAGAGTCAAAAGCTTTCTGTGAAGACAGTCTTCACATCTTGTTCATA 1646
 QY 1081 AACCTAGAGAGAGACAGCAATGATAAAAACAGATGCTAGGAGAGAGACCACTGG 1140
 1647 AACCTAGAGAGAGAGACAGCAATGATAAAAACAGATGCTAGGAGAGAGACCACTGG 1706
 QY 1141 ATGGGCTCACAGACACAGCGCTGAATGAATGAATGAATGAATGAATGAATGAATGAAT 1200
 1707 ATGGGCTCACAGACACAGCGCTGAATGAATGAATGAATGAATGAATGAATGAATGAAT 1766
 QY 1201 GACTACAAAATTTGAAAGCTGACAGCGGATGATGAGGATGATGAGGATGAGGATGAGG 1260
 1767 GACTACAAAATTTGAAAGCTGACAGCGGATGATGAGGATGATGAGGATGAGGATGAGG 1826
 QY 1261 GAAGACTGTGCTGGTGTATTTATGCTGGGAGTGAAGATTTCCAAATGTGAAGACGTC 1320
 1827 GAAGACTGTGCTGGTGTATTTATGCTGGGAGTGAAGATTTCCAAATGTGAAGACGTC 1886
 QY 1321 AATTAATCAATTTGGCAAAAAGACAGGAGACGATGCTGATCTCATTTA 1371
 1887 AATTAATCAATTTGGCAAAAAGACAGGAGACGATGCTGATCTCATTTA 1937

RESULT 3
 AB038518
 LOCUS 3058 bp mRNA linear PRI 08-MAR-2001
 DEFINITION Homo sapiens SCRL mRNA for scavenger receptor with C-type lectin
 type 1, complete cds.
 ACCESSION AB038518
 VERSION AB038518.1 GI:13365514
 KEYWORDS
 SOURCE Homo sapiens tissue_l1b:Placenta cDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Nakamura, K., Funakoshi, H., Miyamoto, K., Tokunaga, F., and Nakamura, T.
 TITLE Molecular cloning and functional characterization of a human
 scavenger receptor with C-type lectin (SCRL), a novel member of a
 scavenger receptor family
 JOURNAL Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
 MEDLINE 2 (bases 1 to 3058)
 NAKAMURA, K. and NAKAMURA, T.
 DIRECT SUBMISSION
 Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical
 School, Division of Biochemistry, Biomedical Research Center, 2-2
 Yamadaoka, Suita, Osaka 565-0871, Japan
 (E-mail: knakamura@onbich.med.osaka-u.ac.jp,
 Tel: 81-6-6879-3783 (ex. 3783), Fax: 81-6-6879-3789)
 FEATURES
 source location/Qualifiers
 1..3058
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

Audrey (US) ; Godowski, Paul J. (US) ; Gurney, Austin L. (US) ;
 Hillan, Kenneth J. (US) ; Marsters, Scot A. (US) ; Pan, James (US)
 ; Paoni, Nicholas F. (US) ; Stephan, Jean-Philippe F. (US) ;
 Watanabe, Colin K. (US) ; Williams, P. Mickey (US) ; Wood, William
 I. (US)

FEATURES
 source
 Location/Qualifiers
 1. 2005
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 606 a 493 c 491 g 415 t
 ORIGIN

Query Match 99.0%; Score 1357.4; DB 6; Length 2005;
 Best Local Similarity 99.9%; Pred. No. 3.2e-305;
 Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

1 ATGAACAGCCAGCTCAATTCATTCACAGAGTCAGATGAGAACATCACCACTATCTCTCAA 60
 134 ATGAACAGCCAGCTCAATTCATTCACAGAGTCAGATGAGAACATCACCACTATCTCTCAA 193
 61 GCCAAGAGCAGACCTGAAAGACCTGCAGACTTACCAAGATGCAAGATAGAAACA 120
 134 GCCAAGAGCAGACCTGAAAGACCTGCAGACTTACCAAGATGCAAGATAGAAACA 253
 121 GCCATCAAGTTCAACCACTGAGAGACGCTTCCAGCTCTTGAGACGATATTGTGAC 180
 254 GCCATCAAGTTCAACCACTGAGAGACGCTTCCAGCTCTTGAGACGATATTGTGAC 313
 181 ATCATTAAGCATATATAGTTTACACAGCCCAACCTGCGAGCCTTGACCAATCTAAT 240
 314 ATCATTAAGCATATATAGTTTACACAGCCCAACCTGCGAGCCTTGACCAATCTAAT 373
 241 GAAGTCAGAGCACTTGCACATATACCTTACCAAAACACAGATGATGTGACCTCTG 300
 374 GAAGTCAGAGCACTTGCACATATACCTTACCAAAACACAGATGATGTGACCTCTG 432
 301 AATAATACCTTGCCCAACATCCGTTTGATTCCTTCTCTAGATGCAAGATTTG 360
 433 AATAATACCTTGCCCAACATCCGTTTGATTCCTTCTCTAGATGCAAGATTTG 492
 361 ATGAGGTGAGGTGAGACCTGAGTACCACTTATCTGATTTATGGAAGAAATGAAG 420
 493 ATGAGGTGAGGTGAGACCTGAGTACCACTTATCTGATTTATGGAAGAAATGAAG 552
 421 CTAGTAGACTCCACAGATGGTCACTCATCAAGATTTTACATATCTACAAGTCCACCG 480
 553 CTAGTAGACTCCACAGATGGTCACTCATCAAGATTTTACATATCTACAAGTCCACCG 612
 481 GGCCCCAGGGGTCCAAGAGTACAGAGATCCCAAGGACCCCTGCGCCCAACTGGCAAC 540
 613 GGCCCCAGGGGTCCAAGAGTACAGAGATCCCAAGGACCCCTGCGCCCAACTGGCAAC 672
 541 AAGGACAGAAAGAGAGAGAGGGGAGCTTGCACCACTGCGCCCTGCGGTGAGAGGCG 600
 673 AAGGACAGAAAGAGAGAGAGGGGAGCTTGCACCACTGCGCCCTGCGGTGAGAGGCG 732
 601 CCAATTGACACAGCTGCTCCCGGAGAGCGTGGGGGAAAGAAAGATCTAAGGCTCCAG 660
 733 CCAATTGACACAGCTGCTCCCGGAGAGCGTGGGGGAAAGATCTAAGGCTCCAG 792
 661 GGCCCCAAAGAGCTCCCGTGTTCCTCTGGGAAAGCCCGGCTTCAGGGGCCCAAGTGGGAC 720
 793 GGCCCCAAAGAGCTCCCGTGTTCCTCTGGGAAAGCCCGGCTTCAGGGGCCCAAGTGGGAC 852
 721 CGAGGCCCCCGGGCCCAACAGGAGAGGATCTCCCGGCTTCAGGGGCCCTCTGGAC 780
 853 CGAGGCCCCCGGGCCCAACAGGAGAGGATCTCCCGGCTTCAGGGGCCCTCTGGAC 912
 781 TTCACAGGACCTTCAGGAGACGCTTGGGGAGCTTGGGGTCTCTGAGACTTGGGGAGCTGCCA 840
 913 TTCACAGGACCTTCAGGAGACGCTTGGGGAGCTTGGGGTCTCTGAGACTTGGGGAGCTGCCA 972
 841 GGCCTTGCTGGGGTACCAAGCATGCGCAGGCCCAAGGGGCCCGGCGCTCTCTGGCCCA 900

973 GGGTTCCTGGGGTACACAGCATGCGACGGCCCAAGAGGCCCGCCGCTCTGGCCCA 1032
 901 TCAGGAGCGGTGTGTCCTGCGCCCTGCAGATGAGCCACCCCGCAGGAGCAAT 960
 1033 TCAGGAGCGGTGTGTCCTGCGCCCTGCAGATGAGCCACCCCGCAGGAGCAAT 1092
 961 GGCTGCGCGCTCAGTGAAGAACCTTACAGCAAAATGCTACTATTTTCAGTTGAGAA 1020
 1093 AGCTGCGCGCTCAGTGAAGAACCTTACAGCAAAATGCTACTATTTTCAGTTGAGAA 1152
 1021 GAAATTTTGGAGATGCAAGCTTTTCTGGAAGCAAGCTTTCACATCTTTTTCATA 1080
 1153 GAAATTTTGGAGATGCAAGCTTTTCTGGAAGCAAGCTTTCACATCTTTTTCATA 1212
 1081 AACACTAGAGAGAACAGCAATGATATAAAAAACAGATGTGAGGAGAGAGCCACTGG 1140
 1213 AACACTAGAGAGAACAGCAATGATATAAAAAACAGATGTGAGGAGAGAGCCACTGG 1272
 1141 ATCGGCTTCACAGACTCAGAGGTGAATGAATGAAATGAAATGAGTGGATGGACATCTCA 1200
 1273 ATCGGCTTCACAGACTCAGAGGTGAATGAATGAAATGAAATGAGTGGATGGACATCTCA 1332
 1201 GACTACAAAAATTTGGAAGCTGAGACCGCGATTAATCGGGTCAATGCGCATGGCCAGGA 1260
 1333 GACTACAAAAATTTGGAAGCTGAGACCGCGATTAATCGGGTCAATGCGCATGGCCAGGA 1392
 1261 GAAGACTGTGCTGGGTGATTTATGCTGGGCGAGTGAACAGATTTTCAATGTGAACGCTC 1320
 1393 GAAGACTGTGCTGGGTGATTTATGCTGGGCGAGTGAACAGATTTTCAATGTGAACGCTC 1452
 1321 AATACTTATTTTGGAAAAAGACAGAGAGACAGTACTGTCAATCTCATTTA 1371
 1453 AATACTTATTTTGGAAAAAGACAGAGAGACAGTACTGTCAATCTCATTTA 1503

RESULT 5
 AX490920 2005 bp DNA linear PAT 16-AUG-2002
 LOCUS
 DEFINITION Sequence 27 from Patent WO0200690.
 ACCESSION AX490920
 VERSION AX490920.1 GI:22323797
 KEYWORDS
 SOURCE
 ORGANISM

Homo sapiens
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1
 Baker, R.P., Ferrara, N., Gerber, H., Gerlitsen, M.E., Goddard, A.,
 Godowski, P.J., Gurney, A.L., Hillan, K.J., Marsters, S.A., Pan, J.,
 Paoni, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I.
 and Ye, W.
 Compositions and methods for the diagnosis and treatment of
 disorders involving angiogenesis
 Patent: WO 0200690-A 27 03-JAN-2002;
 Genentech, Inc. (US)

FEATURES
 source
 Location/Qualifiers
 1. 2005
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 606 a 493 c 491 g 415 t
 ORIGIN

Query Match 99.0%; Score 1357.4; DB 6; Length 2005;
 Best Local Similarity 99.9%; Pred. No. 3.2e-305;
 Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

1 ATGAACAGCCAGCTCAATTCATTCACAGAGTCAGATGAGAACATCACCACTATCTCTCAA 60
 134 ATGAACAGCCAGCTCAATTCATTCACAGAGTCAGATGAGAACATCACCACTATCTCTCAA 193
 61 GCCAAGAGCAGACCTGAAAGACCTGCAGACTTACCAAGATGCAAGATAGAAACA 120

Db 194 GCCACGAGCAGAACTGGAAGACCTGACAGACTTACACAAAGATGCAGAGATAGACA 253
 QY 121 GCCATCAAGTTCAACCAACTGAGAGAAAGCTTCCAGCTTTTGAGAGCGATATTGTGAA 180
 Db 254 GCCATCAAGTTCAACCAACTGAGAGAAAGCTTCCAGCTTTTGAGAGCGATATTGTGAA 313
 QY 181 ATGATTAGCAATATTCAGATTACAGAGCCACCTGCGGAGCGTGCAGCACAATCTTAAT 240
 Db 314 ATGATTAGCAATATTCAGATTACAGAGCCACCTGCGGAGCGTGCAGCACAATCTTAAT 373
 QY 241 GAAGTCAGAGCAGCTTGACAGATACCTTACCAAAACACAGATGATGACCTCTG 300
 Db 374 GAAGTCAGAGCAGCTTGACAGATACCTTACCAAAACACAGATGATGACCTCTG 432
 QY 301 AATTAATACCTGCGCAACATCCGTTGGATTCTGTTCTCTCAGATGACAAGATTG 360
 Db 433 AATTAATACCTGCGCAACATCCGTTGGATTCTGTTCTCTCAGATGACAAGATTG 492
 QY 361 ATGAGGTCGAGGTTAGACACTGAAAGTACCACTTATCAGTGAATGGAAGAATGAAG 420
 Db 493 ATGAGGTCGAGGTTAGACACTGAAAGTACCACTTATCAGTGAATGGAAGAATGAAG 552
 QY 421 CTGATAGACTCCAAAGCTGCTGACGCTCATCAGATATTTTCAATACACTACAAGGCCAG 480
 Db 553 CTGATAGACTCCAAAGCTGCTGACGCTCATCAGATATTTTCAATACACTACAAGGCCAG 612
 QY 481 GGGCCCAAGGGGTCGAAGAGGTGACAGAGATCCAGAGGACCCCTGCGCCCACTGCGAAC 540
 Db 613 GGGCCCAAGGGGTCGAAGAGGTGACAGAGATCCAGAGGACCCCTGCGCCCACTGCGAAC 672
 QY 541 AAGGAGACAGAAAGAGAGAGAGGGGAGCCTGAGACACTGCGCCCTGCGGTTGAGAGAGGC 600
 Db 673 AAGGAGACAGAAAGAGAGAGAGGGGAGCCTGAGACACTGCGCCCTGCGGTTGAGAGAGGC 732
 QY 601 CCAATTGACACAGTGTGCTCCCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCAG 660
 Db 733 CCAATTGACACAGTGTGCTCCCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCAG 792
 QY 661 GGGCCCAAGGGGTCGAGAGGTGCTCCCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCAG 720
 Db 793 GGGCCCAAGGGGTCGAGAGGTGCTCCCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCAG 852
 QY 721 CCAAGGCCCCCGGGGCCCCACAGGCAAGAGAGGACTCCCGGCTCTCAGGGCTCCTGCGC 780
 Db 853 CCAAGGCCCCCGGGGCCCCACAGGCAAGAGAGGACTCCCGGCTCTCAGGGCTCCTGCGC 912
 QY 781 TTCAGAGGACTTTCAGGAGCAGCTTGGGAGAGCTGGGGTGCCTGAGACTTCGGGAGCTGCCA 840
 Db 913 TTCAGAGGACTTTCAGGAGCAGCTTGGGAGAGCTGGGGTGCCTGAGACTTCGGGAGCTGCCA 972
 QY 841 GGCCTTGCTGGGGTACCAAGGATGCCAGGCCCAAGGGCCCCCGGCTCCTGCGCCA 900
 Db 973 GGCCTTGCTGGGGTACCAAGGATGCCAGGCCCAAGGGCCCCCGGCTCCTGCGCCA 1032
 QY 901 TCAGAGAGCGGTGCTGCGCCCTGCGCTGCAGAAATGAGCCAAACCCCGGACCGGAGGACAT 960
 Db 1033 TCAGAGAGCGGTGCTGCGCCCTGCGCTGCAGAAATGAGCCAAACCCCGGACCGGAGGACAT 1092
 QY 961 GGCCTTGCTGGGGTACCAAGGATGCCAGGCCCAAGGGCCCCCGGCTCCTGCGCCA 1020
 Db 1093 AGCTGCGCGCTCAGTCTGGAAGAACTTCACAGACAATGCTACTATTTTTCAGTTGAGAAA 1152
 QY 1021 GAAATTTTGGAGATGCAAAAGCTTCTGTGAAGACAAAGTCTTCACATCTTGTGTTTCATA 1080
 Db 1153 GAAATTTTGGAGATGCAAAAGCTTCTGTGAAGACAAAGTCTTCACATCTTGTGTTTCATA 1212
 QY 1081 AACACTAGAGAGAGACAGCATGATTAATAAACAGATGCTAAGGAGAGAGAGAGAGAGAGAG 1140
 Db 1213 AACACTAGAGAGAGACAGCATGATTAATAAACAGATGCTAAGGAGAGAGAGAGAGAGAGAG 1272
 QY 1141 ATGCGGCTCAGAGACTCAGAGCGGAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 1200
 Db 1273 ATGCGGCTCAGAGACTCAGAGCGGAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 1332

QY 1201 GACTACAAAATTGGAAGCTGGACAGCCGAGTAACTGGGGTCTATGGCCATGGCCAGGA 1260
 Db 1333 GACTACAAAATTGGAAGCTGGACAGCCGAGTAACTGGGGTCTATGGCCATGGCCAGGA 1392
 QY 1261 GAAGACTGCTGGGTTGATTATGCTGGGACAGTGGACATTTCCAAATGTGAAGAGCT 1320
 Db 1393 GAAGACTGCTGGGTTGATTATGCTGGGACAGTGGACATTTCCAAATGTGAAGAGCT 1452
 QY 1321 AATACTTCATTTGCGAAAAAGACAGAGAGAGAGTACTGTCATCTCATTA 1371
 Db 1453 AATACTTCATTTGCGAAAAAGACAGAGAGAGAGTACTGTCATCTCATTA 1503

RESULT 6
 HSN802985
 LOCUS
 DEFINITION Homo sapiens mRNA; cDNA DKFZp547G1215 (from clone DKFZp547G1215).
 ACCESSION AL713657
 VERSION AL713657.1 GI:19584339
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 1886)
 Koehler, K., Beyer, A., Mewes, H.W., Well, B. and Wiemann, S.
 Direct Submission
 Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY

COMMENT
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical Research Center at the Charité, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
 This clone (DKFZp547G1215) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de. Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cdna/>.
 Location/Qualifiers

FEATURES
 source
 1..1886
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="18p11.3"
 /clone="DKFZp547G1215"
 /tissue_type="brain"
 /clone_lib="547 (synonym: hfbp1). Vector psport1; host DH10B; sites NotI + SalI"
 /dev_stage="fetal"
 1..1886
 /gene="DKFZp547G1215"
 <1..1256
 /gene="DKFZp547G1215"
 /note="similarity to asialoglycoprotein receptor (Mus musculus)"

CDS
 gene
 1..1886
 /product="hypothetical protein"
 /protein_id="CAD28466.1"
 /db_xref="GI:19584340"
 /translation="AIFKQLEPEROLEFTDLYVNIISNSTAHHLRLTSLNLEVRT
 TCTDYLTKHTDLSLNNLANIRUDSVSLRMQODIMSRDLTEVAANSLVIEEKL
 DSKHQLIKNFTILLOGPPGRPRGRDGSQGPPTGNKGQGEKGEPPGPAGERG
 PIGPAPGERSGKSGSGQKSGSRSGPQSGSSDPGPAPGPKGKGLPQGP
 PGFQGLQGVGPGPLGPGGLGVPMPQPKPPGPPGPGSAVPLALQNEPTPA
 PPDNCPKMKNFPTKCYFSEYKEKLFEDAKIFCDKSKSHLYFITRQEQMKRQW
 GRESHWIGLIDSERENEMKWDGTSPDKMKRAGQPDMMGHGHEGDEAGLITAGQW
 NDFQEDVNNFICEKDRLETIVSSAL"
 1838..1843
 /gene="DKFZp547G1215"
 1869
 polyA_signal
 polyA_site
 /gene="DKFZp547G1215"

BASE COUNT 569 a 444 c 445 g 428 t

Query Match 79.2%; Score 1086.2; DB 10; Length 2637;
 Best Local Similarity 87.0%; Pred. No. 4.2e-242;
 Matches 1193; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 1 ATGAACGCGCAGCTCACTATTCACAGGTGAGAGAACATCACCCTATCTCTCA 60
 DB 947 ATGAATAGCCAGCTCACTATTCACAGGTGAGAGAACATCACCCTATCTCTCA 1006
 QY 61 GCCACGAGCAGAGAACCTGAGAGACCTGACAGCTTACACAAGAGAGAGAAATAGACA 120
 DB 1007 GCCACGAGCAGAGAACCTGAGAGACCTGACAGCTTACACAAGAGAGAGAAATAGACA 1066
 QY 121 GCCATCAAGTTTCAACCACTGAGAGAACGCTTCAGCTTTTGAAGAGATTTGTAAC 180
 DB 1067 GCTGTCAAGTTTCAAGCTTGAAGAACGCTTCAGCTTTTGAAGAGATTTGTAAC 1126
 QY 181 ATCATTTAGCAATATCACTTACAGCCACCTGCGGAGCTGACACATCTTAAT 240
 DB 1127 ATCATTTAGCAATATCACTTACAGCCACCTGCGGAGCTGACACATCTTAAT 1186
 QY 241 GAAGTCGAGACCACTGACAGATACCTTACCAACACAGATGATGACTGCTT 300
 DB 1187 GATGTTAGAGACCACTGACAGATACCTTACCAACACAGATGATGACTGCTT 1246
 QY 301 AATTAATACCTTGCCCAACATCCGTTTGGATTCGTTTCTCTCAGAGATGACAAGATTTG 360
 DB 1247 AATTAATACCTTGCCCAACATCCGTTTGGATTCGTTTCTCTCAGAGATGACAAGATTTG 1306
 QY 361 ATGAGGCGAGGTTAGACACTGAAGTACCACTTATCACTGATTTAGAGAAATAGAG 420
 DB 1307 ATGAGGCGAGGTTAGACACTGAAGTACCACTTATCACTGATTTAGAGAAATAGAG 1366
 QY 421 CTAGTAGACTCAAGACTGCTGAGCTCATCAAGATTTTCAATTAATCAAGAGTCCACCG 480
 DB 1367 CTGTTGACTCAAGAGCGGTCAAGCTCATCAAGATTTTCAATTAATCAAGAGTCCCT 1426
 QY 481 GGGCCCGAGGGTCCAGAGAGTGCACAGAGATCCAGAGGAGCCCTGCGCCCACTGCGAAC 540
 DB 1427 GGGCCCGAGGGTCCAGAGAGTGCACAGAGATCCAGAGGAGCCCTGCGCCCACTGCGAAC 1486
 QY 541 AAGGACAGAAAGAGAGAGAGAGGAGGAGCTGAGACACTGCGCTGCGGGTGAAGAGGCG 600
 DB 1487 AAGGACAGAAAGAGAGAGAGAGGAGGAGCTGAGACACTGCGCTGCGGGTGAAGAGGCG 1546
 QY 601 CCAATGAGACCTGCTGCTCCCGGAGAGCGTGGCGCAAGAGATCAAGAGCTCCAG 660
 DB 1547 ACAATGAGACCTGCTGCTCCCGGAGAGCGTGGCGCAAGAGATCAAGAGCTCCAG 1606
 QY 661 GGGCCCAAGAGCTGCTGCTCCCGGAGAGCGGCGCTCAAGGCGCCCACTGCGGAG 720
 DB 1607 GGGCCCAAGAGCTGCTGCTCCCGGAGAGCGGCGCTCAAGGAGCTGCGGAG 1666
 QY 721 CCAAGGCCCCGGGCCCCAGAGCAAGAGAGACTCCCGGCGCTCAAGGCGCTCCAG 780
 DB 1667 CCAAGGCCCCGGGCCCCAGAGCAAGAGAGACTCCCGGCGCTCAAGGCGCTCCAG 1726
 QY 781 TTCAGGAGCTTCAAGGAGCTGTTGGGAGCTGGGGTCTGAGACCTGCGGAGCTGCCA 840
 DB 1727 TTCAGGAGCTTCAAGGAGCTGTTGGGAGCTGGGGTCTGAGACCTGCGGAGCTGCCA 1786
 QY 841 GGGTGGCTGGGTACAGAGCAATGCCAGGCCCAAGGCGCCCGGCGCTCCAGGCCCA 900
 DB 1787 GGGTGGCTGGGTACAGAGCAATGCCAGGCCCAAGGCGCCCGGCGCTCCAGGCCCA 1846
 QY 901 TCAAGAGCGGTGGTCCCTGCGCTGAGAGTGAAGCAAGCCCGGAGCGAGAGCAAT 960
 DB 1847 TCAAGAGCGGTGGTCCCTGCGCTGAGAGTGAAGCAAGCCCGGAGCGAGAGCAAT 1906
 QY 961 GGGTGGCGGCTCACTGAGAGAACTTCAAGACAAATGCTACTATTTTCAAGTTGAGAAA 1020
 DB 1907 GGGTGGCGGCTCACTGAGAGAACTTCAAGACAAATGCTACTATTTTCAAGTTGAGAAA 1966
 QY 1021 GAAATTTTGAAGATGCAAGCTTTTCTGTGAAGACAAAGTCTTCAATCTTTGTTTCATA 1080

DB 1967 GAAATTTTGAAGATGCTAGCTTTTCTGTGAAGACAAATCTTCCATCTCTTTTCATA 2026
 QY 1081 AACACTAGAGAGACAGCAATGATTAATAAACACATGCTAGGAGAGAGAGCCACTGG 1140
 DB 2027 AACCTAAGAGAGACAGCAATGATTAATAAACACATGCTAGGAGAGAGAGCCACTGG 2086
 QY 1141 ATCGGCTTCACAGACTCAGAGCGTGAATAATGAATGAGAGCGGTGATGAGATCTCCA 1200
 DB 2087 ATCGGCTTCACAGACTCAGAGCGTGAATAATGAATGAGAGCGGTGATGAGATCTCTT 2146
 QY 1201 GACTACAAAAATTTGAAAGCTGACAGCCGAGTAACTGAGGTCATGCGCCAGCA 1260
 DB 2147 GATTACAAAAATCTGAAAGCTGACCAACAGTAACTGAGGCGAGTGCATGCGCCAGCA 2206
 QY 1261 GAAGACTGCTGCTGGTGAATTTATGCTGGGCGAGTGAAGCAATTTCCAACTGAAAGCTC 1320
 DB 2207 GAAGACTGCTGCTGGTGAATTTATGCTGGGCGAGTGAAGCAATTTCCAACTGAAAGCTC 2266
 QY 1321 AATACTTCAATTTGCGAAAAAGACAGGAGAGAGTACTGTCATCTCATTTA 1371
 DB 2267 AATACTTCAATTTGCTGAAGAGAAAGGAGAGAGTACTGTCATCTCATTTA 2317

RESULT 8
 AB038519 3291 bp mRNA linear ROD 18-JAN-2002
 LOCUS Mus musculus src1 mRNA for scavenger receptor with C-type lectin, complete cds.
 ACCESSION AB038519
 VERSION AB038519.1 GI:18146951
 KEYWORDS
 SOURCE Mus musculus
 ORGANISM Mus musculus CDNA to mRNA.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 1 Nakamura, K., Funakoshi, H., Tokunaga, F., and Nakamura, T. Molecular cloning of a mouse scavenger receptor with C-type lectin (SRC1), a novel member of the scavenger receptor family Blochlin. Biophys. Acta 1522 (1), 53-58 (2001)
 JOURNAL MEDLINE 21575692
 MEDLINE 2 (bases 1 to 3291)
 REFERENCE Nakamura, K. and Nakamura, T.
 AUTHORS Direct Submission
 TITLE Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical School, Division of Biochemistry, Biomedical Research Center; 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
 (E-mail: knakamura@bioch.med.osaka-u.ac.jp, tel:81-6-6879-3783(ex.3783), fax:81-6-6879-3789)
 FEATURES
 source
 1..3291
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 1..3291
 /gene="src1"
 /gene="src1"
 77..2305
 /codon_start=1
 /product="scavenger receptor with C-type lectin"
 /protein_id="BAB82497.1"
 /db_xref="GI:18146952"
 /translation="MKDFAEEEEVQSFGRKRGTHGCTQKCIINMALKRSIVLY ILCAITIVAILGYKVEKMDNVSDGMEHTSHQYDNKLTJAVESDLKLGIDAGKAL STNSELSTFRSDILDLRODLOETETSKNKTLLKLDANGSLVDROSOLKETLOIN SPLITVNTKTOAYNGYVNILOODNTNVOYLOOSMYSQSVINMLNINLNTLOVOOR LLSNLOQSVDTSLAIORIKNDFOMLQOYFDLAKDTPMLKRVSLQTLAANNSALA KANPDLMDNSQLSFTGQMDNITITISQANQSLKDLDLKHTEENRVAVESOLEE RPOVEETDVNTISINISYAHHLRLTSLNDVMTCTDTLRHRDLDLTLNLTVNI RLDLSLRMOQDMRSKLDTEEVANLSVMEEMKLVDSKHGDLIKNFTILQGPGRGP KGRDSOGPPGPTGNKGKGEKGEPPGAPGRTIGVPGGSGSGSGSGSGSGSGPK GSRGSPKRGPGSGGPPGPPGPKDGLPGGPGPPGPGPGGLOGTGEGCVPRPLPG LPGVPMQPPKPPPPGPPGSMPELALQNEPTTASEVNGCPPHMKFTDKCYFSLE

BASE COUNT 1186 a 1053 g 1044 t
ORIGIN

| Query Match | 70.1%; | Score 960.4; | DB 9; | Length 4330; |
|-----------------------|--|---------------------|-----------|--------------|
| Best Local Similarity | 99.9%; | Pred. No. 8.2e-213; | | |
| Matches 961; | Conservative 0; | Mismatches 1; | Indels 0; | Gaps 0; |
| QY 1 | ATGAACACCCAGCTCACTCACTTCACAGGTAGATGAGAACATCACCATCTCTCAA | 60 | | |
| Db 858 | ATGAACACCCAGCTCACTCACTTCACAGGTAGATGAGAACATCACCATCTCTCAA | 917 | | |
| QY 61 | GCCAAACGAGCAACCTTGAAGACCTGCAGAGCTTACACAAAGATGCGAGATATGAACA | 120 | | |
| Db 918 | GCCAAACGAGCAACCTTGAAGACCTGCAGAGCTTACACAAAGATGCGAGATATGAACA | 977 | | |
| QY 121 | GCCATCAGTTCAACCACTGGAGGAGCGTTCCAGCTCTTTGAGACGATATTTGTGAC | 180 | | |
| Db 978 | GCCATCAGTTCAACCACTGGAGGAGCGTTCCAGCTCTTTGAGACGATATTTGTGAC | 1037 | | |
| QY 181 | ATCATTTAGCAATATCATGTTACACACGCCACCACTGCGGAGCGCTGACAGCAATCTAAT | 240 | | |
| Db 1038 | ATCATTTAGCAATATCATGTTACACACGCCACCACTGCGGAGCGCTGACAGCAATCTAAT | 1097 | | |
| QY 241 | GAAAGTCAGGACCACTTGCACAGATACCTTACCAAAACACACAGATGATCTGACCTCTTG | 300 | | |
| Db 1098 | GAAAGTCAGGACCACTTGCACAGATACCTTACCAAAACACACAGATGATCTGACCTCTTG | 1157 | | |
| QY 301 | AATATATACCTTGGCCAAATCCGTTTGGATTTCTGTTCTCAGAGATGCAACAAATTTG | 360 | | |
| Db 1158 | AATATATACCTTGGCCAAATCCGTTTGGATTTCTGTTCTCAGAGATGCAACAAATTTG | 1217 | | |
| QY 361 | ATGAGTGGAGTTAGACACTGTAAGTACCAACTATTCAGTATTTAGGAAGAAATGAAG | 420 | | |
| Db 1218 | ATGAGTGGAGTTAGACACTGTAAGTACCAACTATTCAGTATTTAGGAAGAAATGAAG | 1277 | | |
| QY 421 | CTAGTAGACTCCAGCATGTCTAGCTCATCAAGAATTTTACAAATCTCAAAAGTCCACCG | 480 | | |
| Db 1278 | CTAGTAGACTCCAGCATGTCTAGCTCATCAAGAATTTTACAAATCTCAAAAGTCCACCG | 1337 | | |
| QY 481 | GGCCCCAGGGGTCCAAGAGGTGACAGAGAGATCCCGAGGAGCCCTGGGCCCAACTGGCAAC | 540 | | |
| Db 1338 | GGCCCCAGGGGTCCAAGAGGTGACAGAGAGATCCCGAGGAGCCCTGGGCCCAACTGGCAAC | 1397 | | |
| QY 541 | AAGGACAGAAAGAGAGAGAGGGGAGCGCTGGACCACTGGCCCTGCGGGTAGAGAGAGC | 600 | | |
| Db 1398 | AAGGACAGAAAGAGAGAGAGGGGAGCGCTGGACCACTGGCCCTGCGGGTAGAGAGAGC | 1457 | | |
| QY 601 | CCAATTGACACCAAGCTGTGTCCTCCGAGAGAGCGTGGCGCAAAAGGATCTTAAAGCTCCAG | 660 | | |
| Db 1458 | CCAATTGACACCAAGCTGTGTCCTCCGAGAGAGCGTGGCGCAAAAGGATCTTAAAGGCTCCAG | 1517 | | |
| QY 661 | GGCCCCAAAGGCTCCCGGTGTCCCTCTGGGAAAGCCCGGCTCTCAGGGGCCCAAGTGGGAGC | 720 | | |
| Db 1518 | GGCCCCAAAGGCTCCCGGTGTCCCTCTGGGAAAGCCCGGCTCTCAGGGGCCCAAGTGGGAGC | 1577 | | |
| QY 721 | CCAGGCCCCCGGGGCGCCACAGGCAAAAGGAGACTCCCGGCGCTCAGGGGCTCTCTGGC | 780 | | |
| Db 1578 | CCAGGCCCCCGGGGCGCCACAGGCAAAAGGAGACTCCCGGCGCTCAGGGGCTCTCTGGC | 1637 | | |
| QY 781 | TTCCAGGAGCTTCAGGGGACCGTTGGGGAGCGCTGGGGTCTCTGAGACTCTGGGGAGCTGCA | 840 | | |
| Db 1638 | TTCCAGGAGCTTCAGGGGACCGCGTTGGGGAGCGCTGGGGTCTCTGAGACTCTGGGGAGCTGCA | 1697 | | |
| QY 841 | GCGTTCCTGGGGTACAGGCAATGCCAGGCCCCCAAGGGGCCCTCCCGGCTCTGAGCCCA | 900 | | |

| Accession | Sequence | Position |
|-----------|--|----------|
| Db | 1698 GGCCTTGCCCTGGGGGTCACAGGCATGCGCAGGCCCCCAAGGGCCCCCGGGCCCTCTGGCCCA | 1757 |
| Qy | 901 TCAGAGCGCGGTGTGTCCTGGCCCTGCAGATGAGCCCAACCCGCGACCGAGGACAAAT | 960 |
| Db | 1758 TCAGAGCGCGGTGTGTCCTGGCCCTGCAGATGAGCCCAACCCGCGACCGAGGACAAAT | 1817 |
| Qy | 961 GC 962 | |
| Db | 1818 AC 1819 | |

| | |
|------------|--|
| RESULT | 10 |
| LOCUS | AC016128/c |
| DEFINITION | AC016128 Homo sapiens chromosome 18 clone RP11-324G2 map 18, WORKING DRAFT |
| VERSION | AC016128 |
| ACCESSION | AC016128 |
| KEYWORDS | HTG; HTGS_PHASE1; HTGS_DRAFT. |
| SOURCE | Homo sapiens. |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo; |
| AUTHORS | Birren,B., Linton,L., Nusbaum,C. and Lander,E. |
| JOURNAL | Homo sapiens chromosome 18, clone RP11-324G2 |
| TITLE | Unpublished |
| REFERENCE | 2 (bases 1 to 169088) |
| AUTHORS | Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., |

* NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is

|||||
Db 44758 AGGCTCCAGAGCCGCCCAAGAGCTCCCTGGTTCCCTCGAGAGCCGCCCTCAGGGCTC 44699
OY 711 CAGTGGGAGCCAGGCCCCCGGGCCACAGGCAAGAGGAGTCCCGGGCCCTCAGGG 770
Db 44698 CAGTGGGAGCCAGGCCCCCGGGCCACAGGCAAGAGGAGTCCCGGGCCCTCAGGG 44639
OY 771 CCTCTGCTTCAGGAGCTTCAGGGACCGTGGGGAGCTGGGGTGCCTGACCTCG 830
Db 44638 CCTCTGCTTCAGGAGCTTCAGGGACCGTGGGGAGCTGGGGTGCCTGACCTCG 44579
OY 831 GGGACTCCAGAGCTTCCTGGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGGCCC 890
Db 44578 GGGACTCCAGAGCTTCCTGGGGTACCAAGCATGCCAGGGCCCCCGGGCCC 44519
OY 891 TCCGAGCCCATCAGAGCGGTGGTGGCCCTGAGCCCTCAGAAATGAGCCCAAGCCCGGAC 950
Db 44518 TCCGAGCCCATCAGAGCGGTGGTGGCCCTGAGCCCTCAGAAATGAGCCCAAGCCCGAC 44459
OY 951 GGAGGACAAATGG 962
Db 44458 GGAGGACAAATGG 44447

RESULT 12
AC024368 71044 bp DNA linear HTG 28-FEB-2000
LOCUS Homo sapiens chromosome 11 clone RP11-179K3 map 11, LOW-PASS
DEFINITION
AC024368 AC024368.1 GI:7108157
SEQUENCE SAMPLING
AC024368 HTG: HTGS_PHASE0.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia: Chordata: Craniata: Vertebrata: Euteleostomi:
1 (bases 1 to 71044) Eutheria: Primates: Catarrhini: Hominoidea: Homo.
TITLE Homo sapiens chromosome 11, clone RP11-179K3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 71044)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavsky, L.,
Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearlano, K., Dewar, K., Diaz, J., Dodge, S., Domino, M., Doyle, M.,
Fenebor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J., C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Landers, T., Laroque, K., Lehoczy, J., Levine, R.,
Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,
McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J.,
Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,
Peterson, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olyar, T. M.,
Petersen, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,
Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and
Zody, M.

TITLE Direct Submission
JOURNAL Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L6145
Center clone name: 179_K3

* NOTE: This record contains 90 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 698: contig of 698 bp in length
* 699 798: gap of 100 bp
* 799 1495: contig of 697 bp in length
* 1496 1595: gap of 100 bp
* 1596 2284: contig of 689 bp in length
* 2285 2384: gap of 100 bp
* 2385 3056: contig of 672 bp in length
* 3057 3156: gap of 100 bp
* 3157 3851: contig of 695 bp in length
* 3852 3951: gap of 100 bp
* 3952 4627: contig of 676 bp in length
* 4628 4727: gap of 100 bp
* 4728 5409: contig of 682 bp in length
* 5410 5509: gap of 100 bp
* 5510 6183: contig of 674 bp in length
* 6184 6283: gap of 100 bp
* 6284 6984: contig of 701 bp in length
* 6985 7084: gap of 100 bp
* 7085 7768: contig of 684 bp in length
* 7769 7868: gap of 100 bp
* 7869 8550: contig of 682 bp in length
* 8551 8650: gap of 100 bp
* 8651 9352: contig of 702 bp in length
* 9353 9452: gap of 100 bp
* 9453 10140: contig of 688 bp in length
* 10141 10240: gap of 100 bp
* 10241 10922: contig of 682 bp in length
* 10923 11022: gap of 100 bp
* 11023 11709: contig of 687 bp in length
* 11710 11809: gap of 100 bp
* 11810 12511: contig of 702 bp in length
* 12512 12611: gap of 100 bp
* 12612 13304: contig of 693 bp in length
* 13305 13404: gap of 100 bp
* 13405 14095: contig of 691 bp in length
* 14096 14195: gap of 100 bp
* 14196 14881: contig of 686 bp in length
* 14882 14981: gap of 100 bp
* 14982 15674: contig of 693 bp in length
* 15675 15774: gap of 100 bp
* 15775 16460: contig of 686 bp in length
* 16461 16560: gap of 100 bp
* 16561 17250: contig of 690 bp in length
* 17251 17350: gap of 100 bp
* 17351 18040: contig of 690 bp in length
* 18041 18140: gap of 100 bp
* 18141 18836: contig of 696 bp in length
* 18837 18936: gap of 100 bp
* 18937 19621: contig of 685 bp in length
* 19622 19721: gap of 100 bp
* 19722 20425: contig of 704 bp in length
* 20426 20525: gap of 100 bp
* 20526 21224: contig of 699 bp in length
* 21225 21324: gap of 100 bp
* 21325 22017: contig of 693 bp in length
* 22018 22117: gap of 100 bp
* 22118 22815: contig of 698 bp in length
* 22816 22915: gap of 100 bp

| | | | |
|-------|--------|------------------|-----------|
| 22916 | 23609: | contlg of 694 bp | in length |
| 22910 | 23709: | gap of 100 bp | in length |
| 23110 | 24402: | contlg of 693 bp | in length |
| 24403 | 24502: | gap of 100 bp | in length |
| 24503 | 25101: | contlg of 689 bp | in length |
| 25192 | 25291: | gap of 100 bp | in length |
| 25292 | 25981: | contlg of 690 bp | in length |
| 25882 | 26081: | gap of 100 bp | in length |
| 26775 | 26774: | contlg of 693 bp | in length |
| 26775 | 26874: | gap of 100 bp | in length |
| 26765 | 27563: | contlg of 689 bp | in length |
| 27664 | 28355: | contlg of 692 bp | in length |
| 28356 | 28455: | gap of 100 bp | in length |
| 28456 | 29145: | contlg of 690 bp | in length |
| 29146 | 29245: | gap of 100 bp | in length |
| 29246 | 29943: | contlg of 698 bp | in length |
| 29944 | 30043: | gap of 100 bp | in length |
| 30044 | 30745: | contlg of 702 bp | in length |
| 30746 | 30845: | gap of 100 bp | in length |
| 30846 | 31555: | contlg of 690 bp | in length |
| 31536 | 31655: | gap of 100 bp | in length |
| 31636 | 32359: | contlg of 694 bp | in length |
| 32330 | 32429: | gap of 100 bp | in length |
| 32330 | 33114: | contlg of 685 bp | in length |
| 33115 | 33214: | gap of 100 bp | in length |
| 33215 | 33898: | contlg of 684 bp | in length |
| 33899 | 33998: | gap of 100 bp | in length |
| 33999 | 34694: | contlg of 696 bp | in length |
| 34695 | 34794: | gap of 100 bp | in length |
| 34795 | 35466: | contlg of 692 bp | in length |
| 35487 | 35586: | gap of 100 bp | in length |
| 35587 | 36292: | contlg of 706 bp | in length |
| 36293 | 36392: | gap of 100 bp | in length |
| 36393 | 37095: | contlg of 703 bp | in length |
| 37096 | 37195: | gap of 100 bp | in length |
| 37196 | 37872: | contlg of 677 bp | in length |
| 37873 | 37972: | gap of 100 bp | in length |
| 37973 | 38663: | contlg of 691 bp | in length |
| 38664 | 38763: | gap of 100 bp | in length |
| 38764 | 39449: | contlg of 686 bp | in length |
| 39450 | 39549: | gap of 100 bp | in length |
| 39550 | 40236: | contlg of 687 bp | in length |
| 40237 | 40336: | gap of 100 bp | in length |
| 40337 | 41025: | contlg of 689 bp | in length |
| 41026 | 41125: | gap of 100 bp | in length |
| 41126 | 41818: | contlg of 693 bp | in length |
| 41819 | 41918: | gap of 100 bp | in length |
| 41919 | 42608: | contlg of 690 bp | in length |
| 42609 | 42708: | gap of 100 bp | in length |
| 42709 | 43399: | contlg of 691 bp | in length |
| 43300 | 43499: | gap of 100 bp | in length |
| 43500 | 44169: | contlg of 670 bp | in length |
| 44170 | 44269: | gap of 100 bp | in length |
| 44270 | 44966: | contlg of 697 bp | in length |
| 44667 | 45066: | gap of 100 bp | in length |
| 45067 | 45750: | contlg of 684 bp | in length |
| 45751 | 45850: | gap of 100 bp | in length |
| 45851 | 46545: | contlg of 695 bp | in length |
| 46546 | 46645: | gap of 100 bp | in length |
| 46646 | 47336: | contlg of 691 bp | in length |
| 47337 | 47436: | gap of 100 bp | in length |
| 47437 | 48133: | contlg of 687 bp | in length |
| 48124 | 48223: | gap of 100 bp | in length |
| 48224 | 48899: | contlg of 676 bp | in length |
| 48900 | 48999: | gap of 100 bp | in length |
| 49000 | 49672: | contlg of 673 bp | in length |
| 49673 | 49772: | gap of 100 bp | in length |
| 49773 | 50465: | contlg of 693 bp | in length |
| 50466 | 50565: | gap of 100 bp | in length |
| 50566 | 51234: | contlg of 689 bp | in length |
| 51235 | 51354: | gap of 100 bp | in length |
| 51355 | 52062: | contlg of 708 bp | in length |

| | | | | |
|---|-------|--------|-----------|------------------|
| * | 52063 | 52162: | gap of | 100 bp |
| * | 52163 | 52284: | contig of | 685 bp in length |
| * | 52848 | 52947: | gap of | 100 bp |
| * | 52948 | 53646: | contig of | 639 bp in length |
| * | 53647 | 53746: | gap of | 100 bp |
| * | 53747 | 54409: | contig of | 663 bp in length |
| * | 54410 | 54509: | gap of | 100 bp |

| | | | | |
|-----------------------|-------|--------------------|----------|--------------|
| Query Match | 35.3% | Score 484.6 | DB 2 | Length 71044 |
| Best Local Similarity | 98.2% | Pred. No. 5.2e-102 | | |
| Matches 490 | 0 | Mismatches 9 | Indels 0 | Gaps 0 |

464 TACTACAAGGTCACCGGGCCAGGGGTCCAAGAGGTGACAGAGGATCCAGGACCC 523

Db 43599 TAGCAAAAGTCCACCGGGCCCAAGGGTGCACAGAGGATCCAGGGACCCC 43658

QY 524 CTGGCCCACTGGCAACCAAGGACAGAAAGGAGAGAGGGGAGCCTGGACCACCTGGCC 583

DB 43059 C1GGCCCAAC1GGCAACAAAGGACAGAAAGAGAGAGAGCC1GGACCAAC1GGCC 43/18

[illegible][illegible][illegible]

704 AGGGCCCCAGTTGGGGACCCAGGCCCCCGGGCCACCCAGGCCAAGAGGGACTCCCGGGCC 763

Db 43839 AGGCTCCAGTGGGACCCAGGCCCGGGCCACCAAGGCAAGAGGACTCCCCGGCC 438

QY 764 CTCAGGGCCCTCCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGAGCCCTGGGGTGCCCTG 823

Db 43899 CTCAGGGCCCTCCTGGCTTCCAGGACTTCAGGGCACCGTTGGGAGCCCTGGGGTGCCCTG 439

824 GACCTGGGGACTGCCAGGCTTGCCCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCC 883

Db 43959 GACCTCGGGACTGCCAGGCTTGCTGGGTACCAAGCATGCCAGGCCCAAGGGCCCC 44018

884 CCGCCCCCICCGCCCAICAGGAGCGGIGGIGCCCCCICGCGAAIGAGCCACCC 943

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 104

[illegible]

RESULT 13
AP000900

| | | | | | |
|------------|--|-----------|-----|--------|--------------|
| LOCUS | AP000900 | 1/8022 bp | DNA | linear | HTG 30-MAY-2 |
| DEFINITION | Homo sapiens chromosome 18 clone RP11-683J11 map 18p11.3, WORKIN | | | | |

UNAFI SEQUENCE, 20 UNORDERED PIECES.

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

[illegible]

REFERENCE
1 (bases 1 to 178022)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.

THE
HOMO SAPIENS 1/8,022 GENOMIC DNA CLIP1.1.3
JOURNAL
Published Only in DataBase (1999)

AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,

| TITLE | Direct Submission | Submitted (17 DEC 1980) | Year/Issue | Author | Institution |
|------------|-------------------|-------------------------|------------|--------|-------------|
| TOURNAMENT | | | | | |

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ. 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555.

COMMENT

Japan (E-mail:hatori@gsr.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6997403.

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hatori@gsr.riken.go.jp

----- Project Information

Center project name: HumDraff18

Center clone name: RP11-683J11

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 161944 bases at least Q40
Consensus quality: 170357 bases at least Q40
Consensus quality: 174322 bases at least Q20
Insert size: 176122; sum-of-ctrls
Quality coverage: 4.56x in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 24283 contig of 24283 bp in length
24384 46875 contig of 22492 bp in length
46976 65058 contig of 18083 bp in length
65159 81322 contig of 16164 bp in length
81423 98823 contig of 17401 bp in length
98924 110068 contig of 11145 bp in length
122716 122716 contig of 12548 bp in length
122817 133089 contig of 10273 bp in length
133190 141206 contig of 8017 bp in length
141307 149050 contig of 7744 bp in length
149151 156206 contig of 6955 bp in length
156206 162048 contig of 5843 bp in length
162149 165219 contig of 2970 bp in length
165219 167480 contig of 2262 bp in length
167581 169986 contig of 2406 bp in length
170087 171909 contig of 1823 bp in length
172010 172317 contig of 308 bp in length
172418 174663 contig of 2246 bp in length
174764 176271 contig of 1508 bp in length
176372 178022 contig of 1651 bp in length

Sequence updated (24-Dec-1999)
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 24283: contig of 24283 bp in length
24284 24383: gap of 100 bp
24384 46875: contig of 22492 bp in length
46876 46975: gap of 100 bp
46976 65058: contig of 18083 bp in length
65059 65158: gap of 100 bp
65159 81322: contig of 16164 bp in length
81323 81422: gap of 100 bp
81423 98823: contig of 17401 bp in length
98824 98923: gap of 100 bp
98924 110068: contig of 11145 bp in length
110069 110168: gap of 100 bp
110169 122716: contig of 12548 bp in length
122717 122816: gap of 100 bp

```

```

122817 133089: contig of 10273 bp in length
133090 133189: gap of 100 bp
133190 141206: contig of 8017 bp in length
141207 141306: gap of 100 bp
141307 149050: contig of 7744 bp in length
149051 149150: gap of 100 bp
149151 156105: contig of 6955 bp in length
156106 156205: gap of 100 bp
156206 162048: contig of 5843 bp in length
162049 162148: gap of 100 bp
162149 165118: contig of 2970 bp in length
165119 165218: gap of 100 bp
165219 167480: contig of 2262 bp in length
167481 167580: gap of 100 bp
167581 169986: contig of 2406 bp in length
169987 170086: gap of 100 bp
170087 171909: contig of 1823 bp in length
171910 172009: gap of 100 bp
172010 172317: contig of 308 bp in length
172318 172417: gap of 100 bp
172418 174663: contig of 2246 bp in length
174664 174763: gap of 100 bp
174764 176271: contig of 1508 bp in length
176272 176371: gap of 100 bp
176372 178022: contig of 1651 bp in length.

FEATURES
source
1. 178022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="RP11.3"
/clone="RP11-683J11"
1. 24283
/note="assembly-fragment"
24384. 46875
/note="assembly-fragment"
46976. 65058
/note="assembly-fragment"
65159. 81322
/note="assembly-fragment"
81423. 98823
/note="assembly-fragment"
98924. 110068
/note="assembly-fragment"
110169. 122716
/note="assembly-fragment"
122817. 133089
/note="assembly-fragment"
133190. 141206
/note="assembly-fragment"
141307. 149050
/note="assembly-fragment"
149151. 156105
/note="assembly-fragment"
156206. 162048
/note="assembly-fragment"
162149. 165118
/note="assembly-fragment clone_end:SP6 vector_side:left"
165219. 167480
/note="assembly-fragment"
167581. 169986
/note="assembly-fragment"
170087. 171909
/note="assembly-fragment"
172010. 172317
/note="assembly-fragment clone_end:T7 vector_side:right"
172418. 174663
/note="assembly-fragment"
174764. 176271
/note="assembly-fragment"
176372. 178022
/note="assembly-fragment"
34911 c 36303 g 50935 t 1901 others
BASE COUNT 53972 a

```

ORIGIN

Query Match 34.6%: Score 474; DB 2; Length 178022;

Best Local Similarity 100.0%: Pred. No. 1.6e-99;

Matches 474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGAGACGCCAGCTCACTCATTCACAGGTGAGAGAACATCAACCATATCTCTCAA 60
 DB 13428 ATGAGACGCCAGCTCACTCATTCACAGGTGAGAGAACATCAACCATATCTCTCAA 13487
 OY 61 GCCAAGCAGCAGCACTGAAAGCTGCGAGGACTTACACAAGATGACAGATAGACA 120
 DB 13488 GCCAAGCAGCAGCACTGAAAGCTGCGAGGACTTACACAAGATGACAGATAGACA 13547
 OY 121 GCCATCAAGTTCACCAACCTGAGGAGAGGCTTCTTGAGACGGATATTTGCAAC 180
 DB 13548 GCCATCAAGTTCACCAACCTGAGGAGAGGCTTCTTGAGACGGATATTTGCAAC 13607
 OY 181 ATCATTAGCAATATACATTACACAGCCCAACCTGCGAGGCTGACCAACAATCTAAT 240
 DB 13608 ATCATTAGCAATATACATTACACAGCCCAACCTGCGAGGCTGACCAACAATCTAAT 13667
 OY 241 GAAGTCAGGACCACTTGCACAGATACCTTACCAACACACAGATGATCTGACCTCTTG 300
 DB 13668 GAAGTCAGGACCACTTGCACAGATACCTTACCAACACACAGATGATCTGACCTCTTG 13727
 OY 301 AATATATACCTGCGCAACATCCGTTGGATTCTGTTCTCTCAGATGCAACAAGATTG 360
 DB 13728 AATATATACCTGCGCAACATCCGTTGGATTCTGTTCTCTCAGATGCAACAAGATTG 13787
 OY 361 ATGAGGTCGAGGTAGACACTGAGTACGCACTTATCACTGATTTGGAAGAATGTAAG 420
 DB 13788 ATGAGGTCGAGGTAGACACTGAGTACGCACTTATCACTGATTTGGAAGAATGTAAG 13847
 OY 421 CTAGTGACTCCAGCATGCTGCTCATCAACAATTTTCAATFACTACAGGT 474
 DB 13848 CTAGTGACTCCAGCATGCTGCTCATCAACAATTTTCAATFACTACAGGT 13901

RESULT 14

AP001022

LOCUS 187635 bp DNA linear HTG 30-MAY-2000

DEFINITION

Homo sapiens chromosome 18 clone RP11-815L4 map 18p11.3, WORKING

ACCESSION

AP001022

VERSION

AP001022.2 GI:8117692

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

REFERENCE

Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.

AUTHORS

Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,

TITLE

Fujiyama, A., Yada, T., Totsuki, Y., Watanabe, H. and Sakaki, Y.

JOURNAL

Published Only in Database (2000)

REFERENCE

Submitted (05-JAN-2000) Masahira Hattori, The Institute of Physical

AUTHORS

Kitsato Univ., 1-15-1 Kitsato, Sagami-hara, Kanagawa 228-8555,

COMMENT

URL: http://hgp.gsc.riken.go.jp/; Tel: 81-42-778-9923.

JOURNAL

On May 31, 2000 this sequence version replaced gi:5997772.

CENTER

Center: RIKEN Genomic Sciences Center (GSC)

CONTACT

Web site: http://hgp.gsc.riken.go.jp/

PROJECT

Contact: hattori@gs.riken.go.jp

Center project name: HumDrafl18
 Center clone name: RP11-815L4

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 167192 bases at least Q40

Consensus quality: 177020 bases at least Q30

Consensus quality: 182105 bases at least Q20

Insert size: 184435; sum-of-contigs

Quality coverage: 4.35x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

| | | | | | | |
|--------|--------|-----------|-------|----|----|--------|
| 1 | 21332 | contig of | 21332 | bp | in | length |
| 21433 | 42072 | contig of | 20640 | bp | in | length |
| 42173 | 54336 | contig of | 12164 | bp | in | length |
| 54437 | 66766 | contig of | 12330 | bp | in | length |
| 66867 | 77849 | contig of | 10983 | bp | in | length |
| 77950 | 87798 | contig of | 9849 | bp | in | length |
| 87899 | 99510 | contig of | 11612 | bp | in | length |
| 99611 | 107273 | contig of | 7663 | bp | in | length |
| 107374 | 113189 | contig of | 5816 | bp | in | length |
| 113290 | 120559 | contig of | 7270 | bp | in | length |
| 120660 | 126468 | contig of | 5809 | bp | in | length |
| 126569 | 131619 | contig of | 5051 | bp | in | length |
| 131720 | 136810 | contig of | 5091 | bp | in | length |
| 136911 | 139925 | contig of | 3015 | bp | in | length |
| 140026 | 144232 | contig of | 4207 | bp | in | length |
| 144333 | 148302 | contig of | 4570 | bp | in | length |
| 149003 | 152918 | contig of | 3916 | bp | in | length |
| 153019 | 156399 | contig of | 3381 | bp | in | length |
| 156500 | 158933 | contig of | 2434 | bp | in | length |
| 159034 | 161884 | contig of | 2851 | bp | in | length |
| 161985 | 164301 | contig of | 2317 | bp | in | length |
| 164402 | 166300 | contig of | 1899 | bp | in | length |
| 166401 | 168580 | contig of | 2180 | bp | in | length |
| 168681 | 171270 | contig of | 2590 | bp | in | length |
| 171371 | 174092 | contig of | 2722 | bp | in | length |
| 174193 | 176095 | contig of | 1903 | bp | in | length |
| 176196 | 177890 | contig of | 1695 | bp | in | length |
| 177991 | 179786 | contig of | 1796 | bp | in | length |
| 179887 | 181478 | contig of | 1592 | bp | in | length |
| 181579 | 182895 | contig of | 1317 | bp | in | length |
| 182996 | 184841 | contig of | 1846 | bp | in | length |
| 184942 | 186345 | contig of | 1404 | bp | in | length |
| 186446 | 187635 | contig of | 1190 | bp | in | length |

Sequence updated (26-May-2000).

NOTE: This is a 'working draft' sequence. It currently

consists of 33 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 21332 contig of 21332 bp in length

21433 21432: gap of 100 bp

21433 42072: contig of 20640 bp in length

42073 42172: gap of 100 bp

42173 54336: contig of 12164 bp in length

54337 54436: gap of 100 bp

54437 66766: contig of 12330 bp in length

66767 66866: gap of 100 bp

66867 77849: contig of 10983 bp in length

77850 77949: gap of 100 bp

77950 87798: contig of 9849 bp in length

87799 87898: gap of 100 bp

```
* 87899 99510: contig of 11612 bp in length
* 99511 99610: gap of 100 bp
* 99611 107273: contig of 7663 bp in length
* 107274 107374: gap of 100 bp
* 107374 113189: contig of 5816 bp in length
* 113190 113289: gap of 100 bp
* 113290 120559: contig of 7270 bp in length
* 120560 120659: gap of 100 bp
* 120660 126468: contig of 5809 bp in length
* 126469 126568: gap of 100 bp
* 126569 131619: contig of 5051 bp in length
* 131620 131719: gap of 100 bp
* 131720 136810: contig of 5091 bp in length
* 136811 136910: gap of 100 bp
* 136911 139925: contig of 3015 bp in length
* 139926 140025: gap of 100 bp
* 140026 144232: contig of 4207 bp in length
* 144233 144332: gap of 100 bp
* 144333 148902: contig of 4570 bp in length
* 148903 149002: gap of 100 bp
* 149003 152918: contig of 3916 bp in length
* 152919 153018: gap of 100 bp
* 153019 156399: contig of 3381 bp in length
* 156400 156499: gap of 100 bp
* 156500 158933: contig of 2434 bp in length
* 158934 159033: gap of 100 bp
* 159034 161884: contig of 2851 bp in length
* 161885 161984: gap of 100 bp
* 161985 164301: contig of 2317 bp in length
* 164302 164401: gap of 100 bp
* 164402 166300: contig of 1899 bp in length
* 166301 166400: gap of 100 bp
* 166401 168580: contig of 2180 bp in length
* 168581 168680: gap of 100 bp
* 168681 171270: contig of 2590 bp in length
* 171271 171370: gap of 100 bp
* 171371 174092: contig of 2722 bp in length
* 174093 174192: gap of 100 bp
* 174193 176095: contig of 1903 bp in length
* 176096 176195: gap of 100 bp
* 176196 177890: contig of 1695 bp in length
* 177891 177980: gap of 100 bp
* 177981 179786: contig of 1796 bp in length
* 179787 179886: gap of 100 bp
* 179887 181478: contig of 1592 bp in length
* 181479 181578: gap of 100 bp
* 181579 182895: contig of 1317 bp in length
* 182896 182995: gap of 100 bp
* 182996 184841: contig of 1846 bp in length
* 184842 184941: gap of 100 bp
* 184942 186345: contig of 1404 bp in length
* 186346 186445: gap of 100 bp
* 186446 187635: contig of 1190 bp in length.
*
FEATURES
source
1..187635
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-815L4"
1..21332
/misc_feature
/note="assembly_fragment"
21433..42072
/note="assembly_fragment"
42173..54336
/misc_feature
/note="assembly_fragment"
54437..66766
/misc_feature
/note="assembly_fragment"
66867..77849
/misc_feature
/note="assembly_fragment"
77950..87798
/misc_feature
/note="assembly_fragment clone_end:17 vector_side:right"
87899..99510
```

```
/note="assembly_fragment"
99611..107273
/note="assembly_fragment"
107374..113189
/misc_feature
/note="assembly_fragment clone_end:SP6 vector_side:left"
113290..120559
/note="assembly_fragment"
120660..126468
/misc_feature
/note="assembly_fragment"
126569..131619
/misc_feature
/note="assembly_fragment"
131720..136810
/misc_feature
/note="assembly_fragment"
136911..139925
/misc_feature
/note="assembly_fragment"
140026..144232
/misc_feature
/note="assembly_fragment"
144333..148902
/misc_feature
/note="assembly_fragment"
149003..152918
/misc_feature
/note="assembly_fragment"
153019..156399
```

Query Match 34.6%; Score 474; DB 2; Length 187635;

Best Local Similarity 100.0%; Pred. No. 1,6e-99;

Matches 474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGAACAGCCGCTCACTCATTCATCAGAGTCAGATGGAACATCACCATTCTCA 60
Db 58678 ATGAACAGCCGCTCACTCATTCATCAGAGTCAGATGGAACATCACCATTCTCA 58737
QY 61 GCCAAGCAGCAGAACTGAAAGACCTCGAGACTTACACAAAGATGAGAAATAGACA 120
Db 58738 GCCAAGCAGCAGAACTGAAAGACCTCGAGACTTACACAAAGATGAGAAATAGACA 58797
QY 121 GCCATCAAGTTCACCAACTGGAGAACGCTTCCAGCTCTTTGAGACGATTTGTGAC 180
Db 58798 GCCATCAAGTTCACCAACTGGAGAACGCTTCCAGCTCTTTGAGACGATTTGTGAC 58857
QY 181 ATCATTTGCAATATCATCTTACACAGCCACACCTGGGAGCGTACAGCAATCTAA 240
Db 58858 ATCATTTGCAATATCATCTTACACAGCCACACCTGGGAGCGTACAGCAATCTAA 58917
QY 241 GAAGTCAGAGCACCCTTGACAGATACCTTACCAACACACAGATGATCTGACCTTG 300
Db 58918 GAAGTCAGAGCACCCTTGACAGATACCTTACCAACACACAGATGATCTGACCTTG 58977
QY 301 AATATATCCCTGGCCACATCCGTTTGATTCCTTCTCAGATGCAACAAGATTG 360
Db 58978 AATATATCCCTGGCCACATCCGTTTGATTCCTTCTCAGATGCAACAAGATTG 59037
QY 361 ATGAGTCGAGGTTAGACACTGAAGTGAAGCAACTTATCAGTATGATGAAGAAATGA 420
Db 59038 ATGAGTCGAGGTTAGACACTGAAGTGAAGCAACTTATCAGTATGATGAAGAAATGA 59097
QY 421 CTACTAGACTCAAGCATGCTCAGCTCATCAAGAAATTTTACAACTACAAAGCT 474
Db 59098 CTACTAGACTCAAGCATGCTCAGCTCATCAAGAAATTTTACAACTACAAAGCT 59151
RESULT 15
AP000939 188255 bp DNA linear HTG 30-MAY-2000
LOCUS AP000939
DEFINITION Homo sapiens chromosome 18 clone RP11-839023 map 18p11.3, WORKING
DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION AP000939.3 GI:8119080
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens DNA, Clone:RP11-839023.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 188255)
```

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 TITLE Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 JOURNAL Homo sapiens 188,255 genomic DNA of 18p11.3
 REFERENCE Published Only in Database (1999)
 2 (bases 1 to 188255)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 TITLE Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 JOURNAL Direct Submission
 Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 Kitasato Univ., 1-15-1 Kitasato, Sagaminara, Kanagawa 228-8555,
 Japan (E-mail:hattori@gsc.riken.go.jp,
 URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
 Fax:81-42-778-9924)
 On May 31, 2000 this sequence version replaced gi:6997751.

COMMENT

Center: RIKEN Genomic Sciences Center (GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@gsc.riken.go.jp

Project Information

Center project name: Humdraft18

Center clone name: RPI1-839023

Summary Statistics

Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329

Consensus quality: 171068 bases at least Q40

Consensus quality: 182445 bases at least Q20

Insert size: 185155; sum-of-contigs

Quality coverage: 4.41x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
 32 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs 'N', but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence
 as soon as it is available and the accession number will be
 preserved

```

1 16096 contig of 16096 bp in length
16197 33789 contig of 17593 bp in length
33890 51010 contig of 17121 bp in length
51111 63558 contig of 12448 bp in length
63659 75056 contig of 11398 bp in length
75157 84737 contig of 9581 bp in length
84838 90909 contig of 6072 bp in length
91010 98615 contig of 7606 bp in length
106079 113049 contig of 6971 bp in length
113150 119201 contig of 6052 bp in length
119302 126582 contig of 7281 bp in length
126683 132070 contig of 5388 bp in length
132171 137138 contig of 4968 bp in length
137239 143055 contig of 5817 bp in length
143156 148639 contig of 5484 bp in length
148740 153086 contig of 4347 bp in length
153187 158922 contig of 5735 bp in length
159022 163174 contig of 4133 bp in length
163275 167066 contig of 3792 bp in length
167167 170120 contig of 2954 bp in length
170221 172430 contig of 2210 bp in length
172531 175598 contig of 3068 bp in length
175699 177119 contig of 1420 bp in length
177219 178531 contig of 1313 bp in length
178632 179857 contig of 1226 bp in length
179958 181064 contig of 1107 bp in length
181165 182735 contig of 1571 bp in length
182836 183923 contig of 1008 bp in length
184024 185810 contig of 1787 bp in length
185911 187145 contig of 1225 bp in length
187246 188255 contig of 1010 bp in length
Sequence updated (06-Jan-2000)

```

Sequence updated (26-May-2000).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of 'N', but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 16096: contig of 16096 bp in length
16097 16196: gap of 100 bp
16197 33789: contig of 17593 bp in length
33790 33889: gap of 100 bp
33890 51010: contig of 17121 bp in length
51011 51110: gap of 100 bp
51111 63558: contig of 12448 bp in length
63559 63658: gap of 100 bp
63659 75056: contig of 11398 bp in length
75057 75156: gap of 100 bp
75157 84737: contig of 9581 bp in length
84738 84837: gap of 100 bp
84838 90909: contig of 6072 bp in length
90910 91009: gap of 100 bp
91010 98615: contig of 7606 bp in length
98616 98715: gap of 100 bp
98716 105978: contig of 7263 bp in length
105979 106078: gap of 100 bp
106079 113049: contig of 6971 bp in length
113050 113149: gap of 100 bp
113150 119201: contig of 6052 bp in length
119202 119301: gap of 100 bp
119302 126582: contig of 7281 bp in length
126583 126682: gap of 100 bp
126683 132070: contig of 5388 bp in length
132071 132170: gap of 100 bp
132171 137138: contig of 4968 bp in length
137139 137238: gap of 100 bp
137239 143055: contig of 5817 bp in length
143056 143155: gap of 100 bp
143156 148639: contig of 5484 bp in length
148640 148739: gap of 100 bp
148740 153086: contig of 4347 bp in length
153087 153186: gap of 100 bp
153187 158921: contig of 5735 bp in length
158922 159021: gap of 100 bp
159022 163174: contig of 4133 bp in length
163175 163274: gap of 100 bp
163275 167066: contig of 3792 bp in length
167067 167166: gap of 100 bp
167167 170120: contig of 2954 bp in length
170121 170220: gap of 100 bp
170221 172430: contig of 2210 bp in length
172431 172530: gap of 100 bp
172531 175598: contig of 3068 bp in length
175599 175698: gap of 100 bp
175699 177118: contig of 1420 bp in length
177119 177218: gap of 100 bp
177219 178531: contig of 1313 bp in length
178532 178631: gap of 100 bp
178632 179857: contig of 1226 bp in length
179858 179957: gap of 100 bp
179958 181064: contig of 1107 bp in length
181065 181164: gap of 100 bp
181165 182735: contig of 1571 bp in length
182736 182835: gap of 100 bp
182836 183923: contig of 1008 bp in length
183924 184023: gap of 100 bp
184024 185810: contig of 1787 bp in length
185811 185910: gap of 100 bp
185911 187145: contig of 1225 bp in length
187146 187245: gap of 100 bp
187246 188255: contig of 1010 bp in length.
Location/Qualifiers

```

FEATURES

```
source
1. .188255
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-839023"
1. .16096
/note="assembly_fragment"
16197. .33789
/note="assembly_fragment"
33890. .51010
/note="assembly_fragment"
51111. .63558
/note="assembly_fragment"
63659. .75056
/note="assembly_fragment"
75157. .84737
/note="assembly_fragment"
84838. .90909
/note="assembly_fragment"
91010. .98615
/note="assembly_fragment"
98716. .105978
/note="assembly_fragment"
106079. .113049
/note="assembly_fragment"
113150. .119201
/note="assembly_fragment"
119302. .126582
/note="assembly_fragment"
126683. .132070
/note="assembly_fragment"
132171. .137138
/note="assembly_fragment"
137239. .143055
/note="assembly_fragment"
143156. .148639
/note="assembly_fragment"
148740. .153086
/note="assembly_fragment"
153187. .158921
/note="assembly_fragment"
159022. .163174
misc_feature
33.7%; Score 462; DB 2; Length 188255;
Best Local Similarity 99.8%; Pred. No. 1e-96;
Matches 473; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 ATGAACAGCCGCTCACTCATTCACAGCTCAGATGAGAGAACATCACCACTATCTCTCAA 60
Db 102931 ATGAACAGCCGCTCACTCATTCACAGCTCAGATGAGAGAACATCACCACTATCTCTCAA 102990
QY 61 GCCAACGAGCAAACTGAAAGACTGACAGACTTACACAAGATGCAGAAATAGAAC 120
Db 102991 GCCAACGAGCAAACTGAAAGACTGACAGACTTACACAAGATGCAGAAATAGAAC 103050
QY 121 GCCATCAAGTTCAACCACTGAGAGAAAGCTTCCAGCTTTTGACGAGATATTGTGAAC 180
Db 103051 GCCATCAAGTTCAACCACTGAGAGAAAGCTTCCAGCTTTTGACGAGATATTGTGAAC 103110
QY 181 ATCATTTAGCAATATCACTTACACAGCCACCACTGGGAGCGTGACCACTATAAT 240
Db 103111 ATCATTTAGCAATATCACTTACACAGCCACCACTGGGAGCGTGACCACTATAAT 103170
QY 241 GAAGTCAGAGCACTTGACAGATACCTTACAAACACAGATGATGACCTCCTTG 300
Db 103171 GAAGTCAGAGCACTTGACAGATACCTTACAAACACAGATGATGACCTCCTTG 103230
QY 301 AATTAATACCTGGCCAAACATCCGTTGGATTCTGTTCTCTCAGATGCAACAAGATTG 360
Db 103231 AATTAATACCTGGCCAAACATCCGTTT-GATTCTGTTCTCTCAGATGCAACAAGATTG 103289
QY 361 ATGAGGTGAGGTTAGACACTGAAGTACCACTTATCACTGATTTATGGAAGAAATGAAG 420
```

```
|||||
Db 103290 ATGAGGTGAGGTTAGACACTGAAGTACCACTTATCACTGATTTATGGAAGAAATGAAG 103349
QY 421 CTAGTAGACTCCAAAGCATGGTGCAGCTCATCAACAATTTTACATTAAGGT 474
|||||
Db 103350 CTAGTAGACTCCAAAGCATGGTGCAGCTCATCAACAATTTTACATTAAGGT 103403
```

Search completed: March 21, 2003, 06:35:22
Job time : 4123.17 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 00:58:31 ; Search time 272.589 Seconds
(without alignments)
11326.533 Million cell updates/sec

Title: US-09-763-712a-1-copy_325_1695
Perfect score: 1371
Sequence: 1 atgacagccagctcaactc.....cagtactgtcatctgcatcta 1371

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapect 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
N.GeneSeq_101002:*
1: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneSeq/geneSeqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 1371 | 100.0 | 2024 | 21 | AAA07697 |
| 2 | 1371 | 100.0 | 2262 | 24 | ABA97932 |
| 3 | 1371 | 100.0 | 2628 | 22 | AAH43036 |
| 4 | 1369.4 | 99.9 | 2181 | 22 | AAI60628 |
| 5 | 1369.4 | 99.9 | 2641 | 22 | AAC66903 |
| 6 | 1369.4 | 99.9 | 2929 | 20 | ABQ92072 |
| 7 | 1369.4 | 99.9 | 2930 | 24 | AAV55746 |
| 8 | 1358.4 | 99.1 | 2318 | 22 | AAI58842 |
| 9 | 1357.4 | 99.0 | 2005 | 24 | ABL95574 |

| | | | | | | |
|----|--------|------|------|----|-----------|--------------------|
| 10 | 1357.4 | 99.0 | 2005 | 24 | ABL88085 | Human PRO7223 CDNA |
| 11 | 1086.2 | 79.2 | 2637 | 22 | AAH43037 | Nucleotide sequenc |
| 12 | 883.8 | 64.5 | 1521 | 23 | AA571133 | DNA encoding novel |
| 13 | 617 | 45.0 | 2256 | 22 | AAH43054 | Nucleotide sequenc |
| 14 | 235.6 | 17.2 | 3685 | 20 | AA527858 | Human CSF1 protein |
| 15 | 235.6 | 17.2 | 3810 | 20 | AA527856 | Human CSF1 protein |
| 16 | 169.4 | 12.4 | 493 | 23 | AA571131 | DNA encoding novel |
| 17 | 150.2 | 11.0 | 873 | 23 | AA571130 | DNA encoding novel |
| 18 | 150.2 | 11.0 | 1062 | 23 | AA567442 | DNA encoding novel |
| 19 | 146.8 | 10.7 | 1329 | 23 | AA571134 | DNA encoding novel |
| 20 | 143 | 10.4 | 5041 | 22 | ABA17153 | Human nervous syst |
| 21 | 123 | 9.0 | 1877 | 20 | AA527857 | Human CSF2 protein |
| 22 | 121.6 | 8.9 | 756 | 14 | AAQ43034 | Collagen-like poly |
| 23 | 121.6 | 8.9 | 5676 | 15 | AAQ64556 | Collagen-like poly |
| 24 | 121.4 | 8.9 | 8284 | 22 | AAI26526 | Human breast cance |
| 25 | 121.4 | 8.9 | 8284 | 22 | AAI26526 | Human breast cance |
| 26 | 121.4 | 8.9 | 8284 | 22 | AAI26526 | Human breast cance |
| 27 | 121.4 | 8.9 | 8284 | 22 | AAI26526 | Human breast cance |
| 28 | 121.4 | 8.9 | 8284 | 22 | AAI26526 | Human breast cance |
| 29 | 118.4 | 8.6 | 6109 | 23 | AAH78667 | Human COL5A3 CDNA |
| 30 | 118 | 8.6 | 756 | 14 | AAQ43032 | Collagen-like poly |
| 31 | 118 | 8.6 | 756 | 17 | AAI16766 | Collagen-like poly |
| 32 | 117 | 8.5 | 9287 | 24 | AAH64501 | Human benign prost |
| 33 | 115.8 | 8.4 | 4748 | 22 | AAAD06573 | Human alpha1(I) c |
| 34 | 115.2 | 8.4 | 986 | 22 | AAI189468 | Human polynucleoti |
| 35 | 115 | 8.4 | 4270 | 24 | AAI199885 | Mouse ischaemic co |
| 36 | 114.8 | 8.4 | 4816 | 24 | AAH84041 | Human CDNA differe |
| 37 | 114.8 | 8.4 | 5058 | 23 | AA565026 | DNA encoding novel |
| 38 | 114.8 | 8.4 | 6358 | 22 | AAH96826 | Human EST-derived |
| 39 | 114.8 | 8.4 | 6691 | 23 | AA579806 | DNA encoding novel |
| 40 | 114 | 8.3 | 5060 | 24 | AAK92176 | Prostate cancer-as |
| 41 | 113.8 | 8.3 | 506 | 17 | AAI21298 | pUG4-5-CDK-BP CDNA |
| 42 | 113.6 | 8.3 | 3181 | 19 | AAV59358 | Nucleotide sequenc |
| 43 | 113.4 | 8.3 | 4821 | 23 | AA586866 | DNA encoding novel |
| 44 | 113.4 | 8.3 | 6200 | 23 | AAH78668 | Human COL5A3 CDNA |
| 45 | 113.4 | 8.3 | 6368 | 24 | AAH59647 | Novel human coding |

ALIGNMENTS

| | | |
|----------|---|--|
| RESULT 1 | AAA07697 | standard; DNA; 2024 BP. |
| ID | AAA07697 | |
| AC | AAA07697 | |
| XX | | |
| DT | 20-JUN-2000 | (first entry) |
| XX | | |
| DE | Human collectin encoding DNA. | |
| XX | | |
| KW | Collectin; human; antibacterial; antiviral; ds. | |
| OS | Homo sapiens. | |
| XX | | |
| FH | Key | Location/Qualifiers |
| FT | CDS | 55..1698 |
| FT | | /*tag= "a" collectin" |
| FT | misc_feature | 1..738 |
| FT | | /*tag= "b" |
| FT | misc_feature | /*note= "fragment specifically claimed in claim 8" |
| FT | | 55..738 |
| FT | | /*tag= "c" |
| FT | misc_feature | /*note= "fragment specifically claimed in claim 8" |
| FT | | 79..738 |
| FT | misc_feature | /*tag= "d" |
| FT | | /*note= "fragment specifically claimed in claim 8" |
| FT | misc_feature | 325..738 |
| FT | | /*tag= "e" |
| FT | misc_feature | /*note= "fragment specifically claimed in claim 8" |
| FT | | 358..738 |
| FT | misc_feature | /*tag= "f" |

FT misc-feature /note="Fragment specifically claimed in claim 8"
FT 670..1695
FT /*tag= 9
FT /note="Fragment specifically claimed in claim 5"
FT misc-feature 685..738
FT /*tag= h
FT /note="Fragment specifically claimed in claim 7"
FT misc-feature 730..738
FT /*tag= 1
FT /note="Fragment specifically claimed in claim 7"
FT misc-feature 739..1695
FT /*tag= 1
FT /note="Fragment specifically claimed in claim 6"
FT misc-feature 1696..2024
FT /*tag= k
FT /note="Fragment specifically claimed in claim 9"
XX
XX WO200011161-A1.
XX
XX 02-MAR-2000.
XX
XX 24-AUG-1999; 99WO-JP04552.
XX
XX 24-AUG-1998; 98JP-0237611.
XX
XX (FUSO) FUSO PHARM IND LTD.
XX
XX Wakamlya N;
XX
XX WPI: 2000-224696/19.
XX P-PADB: AAY77985.
XX
XX New collectin of human origin having antibacterial and antiviral
XX activity, and gene encoding it useful for production of transgenic
XX animals and of antibodies for screening potential drug molecules -
XX
XX Claim 5-9; Page 87-90; 106pp; Japanese.
XX
XX The invention relates to polynucleotides encoding a new collectin of
XX human origin. The collectin can be used as an antibacterial and antiviral
XX agent and for screening potential drug molecules. The new collectin can
XX be produced by standard recombinant methodology. The present sequence
XX represents a DNA encoding the human collectin.
XX
XX Sequence 2024 BP; 595 A; 524 C; 506 G; 399 T; 0 other;
SQ
Query Match 100.0%; Score 1371; DB 21; Length 2024;
Best Local Similarity 100.0%; Pred No. 0;
Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAACAGCCAGCTCACTGATTCAGAGTCAGATGAGAACATCACCTATCTCTCAA 60
DB 325 ATGAACAGCCAGCTCACTGATTCAGAGTCAGATGAGAACATCACCTATCTCTCAA 384
QY 61 GCCAAGCAGGAGAACCTGGAAGACCTGACAGCTTACACAAAGATGAGAGATAGACA 120
DB 385 GCCAAGCAGGAGAACCTGGAAGACCTGACAGCTTACACAAAGATGAGAGATAGACA 444
QY 121 GCCATCAAGTTCAACCACTGAGAGAACGCTCCAGCTCTTTGAGAGGATATTGTGAAC 180
DB 445 GCCATCAAGTTCAACCACTGAGAGAACGCTCCAGCTCTTTGAGAGGATATTGTGAAC 504
QY 181 ATCATTAAGCAATATCACTTACACAGCCACACCTGCGAGCGTGACCAATCTTAAT 240
DB 505 ATCATTAAGCAATATCACTTACACAGCCACACCTGCGAGCGTGACCAATCTTAAT 564
QY 241 GAAGTCAGGACCACTTGACAGATACCTTTACCAACACAGATGATCTGACCTCTTG 300
DB 565 GAAGTCAGGACCACTTGACAGATACCTTTACCAACACAGATGATCTGACCTCTTG 624
QY 301 AATAATACCTTGCGCAACATCCGTTTGATTTGTTCTCTCAGAGATGACCAAGATTG 360
DB 625 AATAATACCTTGCGCAACATCCGTTTGATTTGTTCTCTCAGAGATGACCAAGATTG 684

QY 361 ATGAGTCGAGGTTAGACACTGAAGTAGCCACTTATCACTGATTAATGAAATGAAG 420
DB 685 ATGAGTCGAGGTTAGACACTGAAGTAGCCACTTATCACTGATTAATGAAATGAAG 744
QY 421 CTAGTAGACTCCAAAGCATGTGTCACTCATCAAGAAATTTTCAATACTACAGGTCCACCG 480
DB 745 CTAGTAGACTCCAAAGCATGTGTCACTCATCAAGAAATTTTCAATACTACAGGTCCACCG 804
QY 481 GGCCCCGAGGTCGAAGAGTGAAGAGATATCCAGAGACCCCTGAGCCCACTGAGCAAC 540
DB 805 GGCCCCGAGGTCGAAGAGTGAAGAGATATCCAGAGACCCCTGAGCCCACTGAGCAAC 864
QY 541 AAGGACAGAAAGAGAGAAAGGAGGAGCCCTGACCACTGAGCCCTGAGGTCGAGAGAGCC 600
DB 865 AAGGACAGAAAGAGAGAAAGGAGGAGCCCTGACCACTGAGCCCTGAGGTCGAGAGAGCC 924
QY 601 CCAATTGAGACCACTGTCCTCCCGGAGAGACGTGGGGGCAAAAGATCTAAAGCTCCACG 660
DB 925 CCAATTGAGACCACTGTCCTCCCGGAGAGACGTGGGGGCAAAAGATCTAAAGCTCCACG 984
QY 661 GGCCCCAAAGGTCCTCGTGTCTCCCTGAGAGACCCGCTGAGGGCCCACTGAGGAGAC 720
DB 985 GGCCCCAAAGGTCCTCGTGTCTCCCTGAGAGACCCGCTGAGGGCCCACTGAGGAGAC 1044
QY 721 CCAAGGCCCCCGGAGCCACAGGCAAAAGAGGACTCCCGGGCTGAGGGCCCTCTGAGC 780
DB 1045 CCAAGGCCCCCGGAGCCACAGGCAAAAGAGGACTCCCGGGCTGAGGGCCCTCTGAGC 1104
QY 781 TTCCAGGAGACTTTCAGGAGACCGTTGGGAGACCGGGTGTGCTGAGCTTGGGAGACTGCCA 840
DB 1105 TTCCAGGAGACTTTCAGGAGACCGTTGGGAGACCGGGTGTGCTGAGCTTGGGAGACTGCCA 1164
QY 841 GGCTTGCTGGGGTACACGAGCATGCCAGGCGCCCAAGGGCCCGCCCTCTGAGCCCA 900
DB 1165 GGCTTGCTGGGGTACACGAGCATGCCAGGCGCCCAAGGGCCCGCCCTCTGAGCCCA 1224
QY 901 TCAGAGAGCGGTGTGCCCTGCGCTGACAGATGAGCCAAACCCGGACCCGAGAGACAAT 960
DB 1225 TCAGAGAGCGGTGTGCCCTGCGCTGACAGATGAGCCAAACCCGGACCCGAGAGACAAT 1284
QY 961 GGCTGCGCGGCTGACAGGAGAACTTCACAGACAATGCTACATTTTTCAGTTGAGGAA 1020
DB 1285 GGCTGCGCGGCTGACAGGAGAACTTCACAGACAATGCTACATTTTTCAGTTGAGGAA 1344
QY 1021 GAAATTTTGTAGAGATGAAAGCTTTTCTGTGAAGACAAAGCTTCACATCTTGTTCATA 1080
DB 1345 GAAATTTTGTAGAGATGAAAGCTTTTCTGTGAAGACAAAGCTTCACATCTTGTTCATA 1404
QY 1081 AACACTAGAGAGAGAACAGCAATGATTAATAAACAGATGTTAGGAGAGAGAGCCACTGG 1140
DB 1405 AACACTAGAGAGAGAACAGCAATGATTAATAAACAGATGTTAGGAGAGAGAGAGCCACTGG 1464
QY 1141 ATGGGCTTCACAGACTCAGAGCGTGAAGAAATGAGTGAAGGCTGAGATGGAGATCTCCA 1200
DB 1465 ATGGGCTTCACAGACTCAGAGCGTGAAGAAATGAGTGAAGGCTGAGATGGAGATCTCCA 1524
QY 1201 GACTACAAAATTTGGAAGCTGAGACAGCCGATTAAGTGGGTCATGGCCATGGGCCAGGA 1260
DB 1525 GACTACAAAATTTGGAAGCTGAGACAGCCGATTAAGTGGGTCATGGCCATGGGCCAGGA 1584
QY 1261 GAAGACTGTGCTGGGTGATTTATGCTGGGCACTGGAAGCATTTTCCAATGTGAAGAGCTG 1320
DB 1585 GAAGACTGTGCTGGGTGATTTATGCTGGGCACTGGAAGCATTTTCCAATGTGAAGAGCTG 1644
QY 1321 AATAACTCAATTTGCCAAAAAGACAGGAGACAGTACTGTCATCTGCATT 1371
DB 1645 AATAACTCAATTTGCCAAAAAGACAGGAGACAGTACTGTCATCTGCATT 1695

RESULT 2
AB97932
ID AB97932 standard; cDNA; 2262 BP.

```
XX ABA97932;
AC
XX
XX 25-APR-2002 (first entry)
DE Human scavenger receptor-like protein encoding cDNA SEQ ID NO 1.
XX
XX Human scavenger receptor; diagnosis; treatment; autoimmune disease;
XX rheumatoid arthritis; gene; ss.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 33..2262
XX FT /*tag=a
XX FT /product="scavenger receptor-like protein"
XX
XX JP2001340089-A.
XX
XX 11-DEC-2001.
XX
XX 08-DEC-2000; 2000JP-0375066.
XX
XX 27-MAR-2000; 2000JP-0090772.
XX
XX (SHIO ) SHIONOGI & CO LTD.
XX
XX WPI: 2002-144965/19.
XX P-PSDB; ABB08642.
XX
XX New scavenger receptor-like protein for diagnosis, prevention and
XX treatment of autoimmune disease, such as rheumatoid arthritis -
XX
XX Claim 3; Fig 1; 38pp; Japanese.
XX
XX The invention relates to a human scavenger receptor-like protein. The
XX protein is useful as a target molecule for diagnosis, prevention and
XX treatment of autoimmune diseases such as rheumatoid arthritis.
XX
XX Sequence 2262 BP; 670 A; 576 C; 581 G; 435 T; 0 other:
XX
Query Match 100.0%; Score 1371; DB 24; Length 2262;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGACAGGCGACGTCACATTCATTCACAGGTCAGATGAGAGATCACCATATCTCTCAA 60
DB 889 ATGAAACAGCGACGTCACATTCATTCACAGGTCAGATGAGAGATCACCATATCTCTCAA 948
QY 61 GCCAAGCAGCAGACCTGAAAGACCTGCAGACTTACACAAAGATGACAGAAATAGACA 120
DB 949 GCCAAGCAGCAGACCTGAAAGACCTGCAGACTTACACAAAGATGACAGAAATAGACA 1008
QY 121 GCCATCAAGTTCAACCACTGAGGAAGCGCTTCACAGCTCTTGAGACGGATATTGTAAC 180
DB 1009 GCCATCAAGTTCAACCACTGAGGAAGCGCTTCACAGCTCTTGAGACGGATATTGTAAC 1068
QY 181 ATCATTAGCAATATGATGATACACAGCCACAGCTGCGGACGCTGACCGCAATCTAAAT 240
DB 1069 ATCATTAGCAATATGATGATACACAGCCACAGCTGCGGACGCTGACCGCAATCTAAAT 1128
QY 241 GAAGTCAGAGCAGCTTGCACAGATACCTTACCAAAACACAGATATGACCTCTTG 300
DB 1129 GAAGTCAGAGCAGCTTGCACAGATACCTTACCAAAACACAGATATGACCTCTTG 1188
QY 301 AATPATACCTTGCCCAACATCCGTTTGATTCCTCTCAGATGCAACAAGATTG 360
DB 1189 AATPATACCTTGCCCAACATCCGTTTGATTCCTCTCAGATGCAACAAGATTG 1248
QY 361 ATGAGGTGAGGTGAGACCTGAGAGGCACTTATCAGTATTTGGAAGAAATGAAG 420
DB 1249 ATGAGGTGAGGTGAGACCTGAGAGGCACTTATCAGTATTTGGAAGAAATGAAG 1308
```

```
QY 421 CTAGTAGACTCCAGACATGCTGACATCATCAAGATTTTACAAATTTACTACAGATCCACCG 480
DB 1309 CTAGTAGACTCCAGACATGCTGACATCATCAAGATTTTACAAATTTACTACAGATCCACCG 1368
QY 481 GGGCCCCAGGGGCTCCAGAGAGGTGACAGAGATCCAGAGAGACCCCTGGCCCACTGGCAAC 540
DB 1369 GGGCCCCAGGGGCTCCAGAGAGGTGACAGAGATCCAGAGAGACCCCTGGCCCACTGGCAAC 1428
QY 541 AAGGACACAAAGAGAGAGAGAGGGGAGCCTGGACACCACTGGCCCTCGGTGAGAGAGC 600
DB 1429 AAGGACACAAAGAGAGAGAGAGGGGAGCCTGGACACCACTGGCCCTCGGTGAGAGAGC 1488
QY 601 CCAATTGAGACAGACCTGCTCCCGGAGAGACGGTGGGGCAAAAGATCTAAAGCTCCAG 660
DB 1489 CCAATTGAGACAGACCTGCTCCCGGAGAGACGGTGGGGCAAAAGATCTAAAGCTCCAG 1548
QY 661 GGGCCCCAAAGGCTCCCGTGGGTTCCCTGGGAGACCCGGCCCTCAGAGGCCCACTGGGAGC 720
DB 1549 GGGCCCCAAAGGCTCCCGTGGGTTCCCTGGGAGAGCCGGCCCTCAGAGGCCCACTGGGAGC 1608
QY 721 CCAGGCCCCCGGGCCACAGGCAAAAGAGGACTCCCGGCTTACAGGAGCCCTCTGGC 780
DB 1609 CCAGGCCCCCGGGCCACAGGCAAAAGAGGACTCCCGGCTTACAGGAGCCCTCTGGC 1668
QY 781 TTCCAGGAGCTTCAGGAGCACCCTTGGGAGACCTGGGGTGCCTGAGACTGAGGAGCTGCA 840
DB 1669 TTCCAGGAGCTTCAGGAGCACCCTTGGGAGACCTGGGGTGCCTGAGACTGAGGAGCTGCA 1728
QY 841 GGGTTGCTGGGGTACAGGACATGCCAGGCCCAAGGGGCCCGGCCCTCTGCGCCA 900
DB 1729 GGGTTGCTGGGGTACAGGACATGCCAGGCCCAAGGGGCCCGGCCCTCTGCGCCA 1788
QY 901 TCAGAGAGCGGTGCTGCCCCCTGGCCCTGCAGAAATGACACCAACCCCGGACCGAGAGCAAT 960
DB 1789 TCAGAGAGCGGTGCTGCCCCCTGGCCCTGCAGAAATGACACCAACCCCGGACCGAGAGCAAT 1848
QY 961 GGGTGGCCCCCTACACGAGAACTTCACAGCAAAATGCTACTATTTTCAGTGAAGAA 1020
DB 1849 GGGTGGCCCCCTACACGAGAACTTCACAGCAAAATGCTACTATTTTCAGTGAAGAA 1908
QY 1021 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGGAAGACAGTCTTCACATCTTGTTCATA 1080
DB 1909 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGGAAGACAGTCTTCACATCTTGTTCATA 1968
QY 1081 AACACTAGAGAGAGAAACAGCAATGATTAATAAAACAGATGATGAGAGAGAGAGCCACTGG 1140
DB 1969 AACACTAGAGAGAGAAACAGCAATGATTAATAAAACAGATGATGAGAGAGAGAGCCACTGG 2028
QY 1141 ATCGGCTCAGACACTCAGAGCGTGAATAATGAATGGAAGTGGCTGATGGAGCATCTCCA 1200
DB 2029 ATCGGCTCAGACACTCAGAGCGTGAATAATGAATGGAAGTGGCTGATGGAGCATCTCCA 2088
QY 1201 GACTACAAAATTTGGAAGCTGAGACGCCGATTAAGTGGGGTCAATGGGCCAGGA 1260
DB 2089 GACTACAAAATTTGGAAGCTGAGACGCCGATTAAGTGGGGTCAATGGGCCAGGA 2148
QY 1261 GAAGACTGTGCTGGTGAATTTATGCTGGGCACTGGAAGATTTCCAAATGTGAAGACGT 1320
DB 2149 GAAGACTGTGCTGGTGAATTTATGCTGGGCACTGGAAGATTTCCAAATGTGAAGACGT 2208
QY 1321 AATAACTTATTTGCGAAAAAGACAGAGGAGACAGTACTGTCAATCTCATTA 1371
DB 2209 AATAACTTATTTGCGAAAAAGACAGAGGAGACAGTACTGTCAATCTCATTA 2259
RESULT 3
AAH43036
ID AAH43036 standard; DNA; 2628 BP.
XX
XX AAH43036;
XX
XX 15-OCT-2001 (first entry)
XX
```

DE Nucleotide sequence of a human scavenger receptor.
XX
KW Human: scavenger receptor; SRCL-PI; macrophage; basal immunity;
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
KW low density lipoprotein; ss.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 74..2302
FT /tag= a
FT /product= "scavenger receptor"
XX
PN WO200159107-A1.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001: 2001WO-JP00874.
XX
PR 14-FEB-2000: 2000JP-0035155.
PR 10-OCT-2000: 2000JP-0309068.
XX
XX (FUSO) FUSO PHARM IND LTD.
XX
PI Wakamiya N:
XX
XX WPI: 2001-497076/54.
DR P-PSDB: AAG63346.
XX
PT New Scavenger receptor proteins SRCL-PI with collectin-like structure,
PT useful for treatment and diagnosis of diseases associated with oxidized
PT low-density lipoprotein accumulation
XX
PS Claim 2; Page 79-84; 118pp; Japanese.
XX
XX The present sequence encodes a human scavenger receptor, designated
CC SRCL-PI. The SRCL-PI polypeptide has a collectin-like structure. They
CC are useful in clarifying the functions of macrophages and basal
CC immunity. They are also useful in the treatment, prevention, diagnosis
CC and investigation of diseases such as arteriosclerosis, diabetic
CC complications, bacterial infection and restenosis following angioplasty,
CC which are associated with accumulation of oxidized low density
CC lipoprotein and the binding of advanced glycation end-products into
CC cells.
XX
SQ Sequence 2628 BP; 783 A; 661 C; 656 G; 528 T; 0 other;
Query Match 100.0%; Score 1371; DB 22; Length 2628;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAACAGCCAGCTCACTCATTCACAGTCAGATGAGAAATCACCACCTATCTCA 60
DB 929 ATGAACAGCCAGCTCACTCATTCACAGTCAGATGAGAAATCACCACCTATCTCA 988
QY 61 GCCAAGCAGCAGAACTGAAAGACTGACGACTTACACAAAGATGAGAAATAGACA 120
DB 989 GCCAAGCAGCAGAACTGAAAGACTGACGACTTACACAAAGATGAGAAATAGACA 1048
QY 121 GCCATCAAGTTCAACCAACGAGGAGAACGCTTCAGCTTTGAGACGATATTGTAAC 180
DB 1049 GCCATCAAGTTCAACCAACGAGGAGAACGCTTCAGCTTTGAGACGATATTGTAAC 1108
QY 181 ATCATTAAGCAATATACATTACAGCCACACCTGGAGCGTGACCAATCTAAAT 240
DB 1109 ATCATTAAGCAATATACATTACAGCCACACCTGGAGCGTGACCAATCTAAAT 1168
QY 241 GAAGTCAGACCACTGTCACAGATACCTTACCAAAACACAGATGATGACCTCCTG 300
DB 1169 GAAGTCAGACCACTGTCACAGATACCTTACCAAAACACAGATGATGACCTCCTG 1228
QY 301 AATAAATACCTGGCCAAATCCGTTGGATTCTGTTCTCTCAGATGCAAAAGATTG 360
|||||

DB 1229 AATAAATACCTGGCCAAATCCGTTGGATTCTGTTCTCTCAGATGCAAAAGATTG 1288
QY 361 ATGAGTCGAGGTTAGACATGAAATAGCCAACTTACATGATTAAGAAAGAAATGAG 420
DB 1289 ATGAGTCGAGGTTAGACATGAAATAGCCAACTTACATGATTAAGAAAGAAATGAG 1348
QY 421 CTAGTAGACTCCAAAGCATGCTCAGCTCATCAACAATTTTACATCTACAGAGTCCACG 480
DB 1349 CTAGTAGACTCCAAAGCATGCTCAGCTCATCAACAATTTTACATCTACAGAGTCCACG 1408
QY 481 GGGCCCAAGGGGTCAGAGGTGACAGAGATGCCAGGAGCCCTGGCCCAACTGGCAAC 540
DB 1409 GGGCCCAAGGGGTCAGAGGTGACAGAGATGCCAGGAGCCCTGGCCCAACTGGCAAC 1468
QY 541 AAGGACAGAAAGAGAGAGAGGGGAGCCCTGAGCACTTGCCCTGGGCTGAGAGAGGC 600
DB 1469 AAGGACAGAAAGAGAGAGAGGGGAGCCCTGAGCACTTGCCCTGGGCTGAGAGAGGC 1528
QY 601 CCAATTGACACAGCTGCTCCCGGAGAGCTGGGCGGCAAGGATCTAAAGCTCCACG 660
DB 1529 CCAATTGACACAGCTGCTCCCGGAGAGCTGGGCGGCAAGGATCTAAAGCTCCACG 1588
QY 661 GGGCCCAAGGGGTCAGAGGTGCTCCCTGGGAGAGCCCGGCTCAGGGCCCTCAGTGGGAGC 720
DB 1589 GGGCCCAAGGGGTCAGAGGTGCTCCCTGGGAGAGCCCGGCTCAGGGCCCTCAGTGGGAGC 1648
QY 721 CCAAGGCCCCCGGGCCACAGGCAAAAGAGGACCTCCCGGCTCAGGGCCCTCCTGGGC 780
DB 1649 CCAAGGCCCCCGGGCCACAGGCAAAAGAGGACCTCCCGGCTCAGGGCCCTCCTGGGC 1708
QY 781 TTCAGGAGCTTCAGGGGACCGTTGGGGAGCTGGGGGCTGAGACCTGGGGAGCTGCCA 840
DB 1709 TTCAGGAGCTTCAGGGGACCGTTGGGGAGCTGGGGGCTGAGACCTGGGGAGCTGCCA 1768
QY 841 GGCCTGCTGGGGTACCAAGCATGCGAGGCGCAAGGGCCCCCGGCTCCTGGGCCCA 900
DB 1769 GGCCTGCTGGGGTACCAAGCATGCGAGGCGCAAGGGCCCCCGGCTCCTGGGCCCA 1828
QY 901 TCAGAGCGGTGTGTCCTCGGCTGAGAAATGAGCAACCCCGGACCGGAGGAGCAAT 960
DB 1829 TCAGAGCGGTGTGTCCTCGGCTGAGAAATGAGCAACCCCGGACCGGAGGAGCAAT 1888
QY 961 GGCCTGCGGCTCAGTGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1020
DB 1889 GGCCTGCGGCTCAGTGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1948
QY 1021 GAAATTTTGGAGATGCAAAAGCTTTTCTGGAAGACAAAGCTTCACATCTGTTTCATA 1080
DB 1949 GAAATTTTGGAGATGCAAAAGCTTTTCTGGAAGACAAAGCTTCACATCTGTTTCATA 2008
QY 1081 AACACTAGAGAGAGACGCAATGATATAAAAAACAGATGTTAGGAGAGAGACCACTGG 1140
DB 2009 AACACTAGAGAGAGACGCAATGATATAAAAAACAGATGTTAGGAGAGAGACCACTGG 2068
QY 1141 ATGGGCTTCACAGACTGAGAGCTGAAATGAATGGAATGGCTGATGGGACATCTCCA 1200
DB 2069 ATGGGCTTCACAGACTGAGAGCTGAAATGAATGGAATGGCTGATGGGACATCTCCA 2128
QY 1201 GACTACAAAATTTGGAAGCTGGACAGCCGATTAAGCTGGGGTATGGCCATGGGCCAGGA 1260
DB 2129 GACTACAAAATTTGGAAGCTGGACAGCCGATTAAGCTGGGGTATGGCCATGGGCCAGGA 2188
QY 1261 GAAGACTGTGCTGGGTTGATTTATGCTGGGCACTGGAACGATTTCCAAATGTGAAGAGCT 1320
DB 2189 GAAGACTGTGCTGGGTTGATTTATGCTGGGCACTGGAACGATTTCCAAATGTGAAGAGCT 2248
QY 1321 AATAAATCTATTTTGGGAAAAAGACAGGAGACAGTACTGTGATCTGCATTTA 1371
DB 2249 AATAAATCTATTTTGGGAAAAAGACAGGAGACAGTACTGTGATCTGCATTTA 2299
RESULT 4
AAT160628

| | | |
|----|---|--|
| XX | ID | AA160628 standard; cDNA; 2181 BP. |
| XX | AC | AA160628; |
| XX | DT | 22-OCT-2001 (first entry) |
| XX | DE | Human polynucleotide SEQ ID NO 4617. |
| XX | XX | |
| KW | Human; noctropic; immunosuppressant; cyostatic; gene therapy; cancer; | |
| KW | peripheral nervous system; neuropathy; central nervous system; CNS; | |
| KW | Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; | |
| KW | amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; | |
| KW | chemokineic; thrombolytic; drug screening; arthritis; inflammation; | |
| KW | leukaemia; ss. | |
| XX | OS | Homo sapiens. |
| XX | PN | WO200153312-A1. |
| XX | PD | 26-JUL-2001. |
| XX | PF | 26-DEC-2000; 2000WO-US34263. |
| XX | PR | 21-JAN-2000; 2000US-0488725. |
| XX | PR | 25-APR-2000; 2000US-0552317. |
| XX | PR | 09-JUL-2000; 2000US-0598042. |
| XX | PR | 13-JUL-2000; 2000US-0620312. |
| XX | PR | 03-AUG-2000; 2000US-0653450. |
| XX | PR | 14-SEP-2000; 2000US-0662191. |
| XX | PR | 19-OCT-2000; 2000US-0693036. |
| XX | PR | 29-NOV-2000; 2000US-0727344. |
| XX | PA | (HXSE-) HXSEQ INC. |
| XX | PI | Tang Y ^T , Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; |
| XX | PI | Wang J, Wang Z, Weinman T, Xu C, Xue AU, Yang Y, Zhang J; |
| XX | PI | Zhao QA, Zhou P, Goodrich R, Drmanac RT; |
| XX | XX | |
| XX | XX | WPI: 2001-442253/47. |
| XX | XX | P-PSDB; AAM41472. |
| XX | PT | Novel nucleic acids and polypeptides, useful for treating disorders |
| XX | PT | such as central nervous system injuries - |
| XX | PS | Claim 1; SEQ ID NO 4617; 10078pp; English. |
| XX | XX | |
| CC | CC | The invention relates to human nucleic acids (AA157798-AA161369) and |
| CC | CC | the encoded polypeptides (AAM38642-AAM42213) with noctropic, |
| CC | CC | immunosuppressant and cyostatic activity. The polynucleotides are useful |
| CC | CC | in gene therapy. A composition containing a polypeptide or polynucleotide |
| CC | CC | of the invention may be used to treat diseases of the peripheral nervous |
| CC | CC | system, such as peripheral nervous injuries, peripheral neuropathy and |
| CC | CC | localised neuropathies and central nervous system diseases, such as |
| CC | CC | Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic |
| CC | CC | lateral sclerosis, and Shy-Drager Syndrome. Other uses include the |
| CC | CC | utilisation of the activities such as: Immune system suppression, |
| CC | CC | Activin/inhibin activity, chemotactic/chemokineic activity, haemostatic |
| CC | CC | and thrombolytic activity, cancer diagnosis and therapy, drug screening, |
| CC | CC | assays for receptor activity, arthritis and inflammation, leukaemias and |
| CC | CC | C.N.S disorders. |
| CC | CC | Note: The sequence data for this patent did not form part of the printed |
| CC | CC | specification. |
| XX | XX | |
| XX | XX | Sequence 2181 BP: 664 A; 539 C; 535 G; 443 T; 0 other; |
| QY | Query Match | 99.9%; Score 1369.4; DB 22; Length 2181; |
| | Best Local Similarity | 99.9%; Pred. No. 0; |
| | Matches 1370; Conservative | 0; Mismatches 1; Indels 0; Gaps 0; |
| DB | 1 | ATGAACAGCAGCAGCTCACTCACTTCACAGGTCAGATGGAGAACATCAGACCACTATCTCTCAA 60 |
| | 317 | ATGAACAGCAGCAGCTCACTTCACAGGTCAGATGGAGAACATCAGACCACTATCTCTCAA 376 |

| | | | |
|----|------|--|------|
| QY | 61 | GCACAGAGAGAAACCTGAAAGCCCTGACGAGCTTACACAAAGATGACAGATTATGAAACA | 120 |
| Db | 377 | GCACAGAGAGAAACCTGAAAGCCCTGACGAGCTTACACAAAGATGACAGATTATGAAACA | 436 |
| QY | 121 | GCACATCAAGTTCAACCAACTGAGAGAAACCTTCCAGACTTTTGAGACGGATATTGTGAAC | 180 |
| Db | 437 | GCACATCAAGTTCAACCAACTGAGAGAAACCTTCCAGACTTTTGAGACGGATATTGTGAAC | 496 |
| QY | 181 | ATCATTTAGCAATATCAGTTTACACAGCCCAACACTGCGAGCGCTGACGACATCTAAAT | 240 |
| Db | 497 | ATCATTTAGCAATATCAGTTTACACAGCCCAACACTGCGAGCGCTGACGACATCTAAAT | 556 |
| QY | 241 | GAATCGAGGACCACTTGACAGATACCCTTTACCAACACACAGATGATCTGACCTCCCTG | 300 |
| Db | 557 | GAATCGAGGACCACTTGACAGATACCCTTTACCAACACACAGATGATCTGACCTCCCTG | 616 |
| QY | 301 | AATATATACCCCTGGCCCAACATCCGTTTGATCTGTTTCTCTAGGATGCACAAAGATTG | 360 |
| Db | 617 | AATATATACCCCTGGCCCAACATCCGTTTGATCTGTTTCTCTAGGATGCACAAAGATTG | 676 |
| QY | 361 | ATGAGTCTGAGGTTTGAACACTGAAAGTACCAACTTATCAGTATTTATGGAAGAAATGAG | 420 |
| Db | 677 | ATGAGTCTGAGGTTTGAACACTGAAAGTACCAACTTATCAGTATTTATGGAAGAAATGAG | 736 |
| QY | 421 | CTACTAATCTCCAAAGCATGCTGACGCTCATCAGAAATTTTTCATATCTACAAAGTCCACCG | 480 |
| Db | 737 | CTACTAATCTCCAAAGCATGCTGACGCTCATCAGAAATTTTTCATATCTACAAAGTCCACCG | 796 |
| QY | 481 | GGCCCCAGGGGTCCAAGAGGTGACAAAGATCCCAAGGACCCCTGCGCCCAACTGGCAAC | 540 |
| Db | 797 | GGCCCCAGGGGTCCAAGAGGTGACAAAGATCCCAAGGACCCCTGCGCCCAACTGGCAAC | 856 |
| QY | 541 | AAGGACAGAAAGAGAGAGAGGGGGAGCCTTGACCACTGGCCCTGCGGGTGAGAGAGGC | 600 |
| Db | 857 | AAGGACAGAAAGAGAGAGAGGGGGAGCCTTGACCACTGGCCCTGCGGGTGAGAGAGGC | 916 |
| QY | 601 | CCATTTGAGACACTGCTGCTCCCCCGGAGAGCGTGGGGGCAAGATTTAAAGCTCCACG | 660 |
| Db | 917 | CCATTTGAGACACTGCTGCTCCCCCGGAGAGCGTGGGGGCAAGATTTAAAGCTCCACG | 976 |
| QY | 661 | GGCCCCAAAGGCTCCGCTGCTTCCCTCTGGAGAACCCGCGCCCTCAGAGGCCCACTGTTGGGAC | 720 |
| Db | 977 | GGCCCCAAAGGCTCCGCTGCTTCCCTCTGGAGAACCCGCGCCCTCAGAGGCCCACTGTTGGGAC | 1036 |
| QY | 721 | CCAGGGCCCCCGGGCCCAACAGGCAAGAGAGGACTCCCGGCTCTAGGGGCCCTCTGGC | 780 |
| Db | 1037 | CCAGGGCCCCCGGGCCCAACAGGCAAGAGAGGACTCCCGGCTCTAGGGGCCCTCTGGC | 1096 |
| QY | 781 | TTTCAGAGGACTTCAAGGGCACCTTTGGGGGAGCCTGGGGTCTCTGACCTCGGGGACTCGCA | 840 |
| Db | 1097 | TTTCAGAGGACTTCAAGGGCACCTTTGGGGGAGCCTGGGGTCTCTGACCTCGGGGACTCGCA | 1156 |
| QY | 841 | GGCTTGGCTGGGGTACCAAGCATGCGCAAGGGCCCCAAGGGCCCCCGGCTCTGGGCCA | 900 |
| Db | 1157 | GGCTTGGCTGGGGTACCAAGCATGCGCAAGGGCCCCAAGGGCCCCCGGCTCTGGGCCA | 1216 |
| QY | 901 | TCAGAGACGGGTGTCCTGTCCTGGCCCTGCGCAATGAGCCCAACCCCGGCAACCGGAGACAAT | 960 |
| Db | 1217 | TCAGAGACGGGTGTCCTGTCCTGGCCCTGCGCAATGAGCCCAACCCCGGCAACCGGAGACAAT | 1276 |
| QY | 961 | GGCTGCCCCCTCACTGAGAGAACTTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA | 1020 |
| Db | 1277 | GGCTGCCCCCTCACTGAGAGAACTTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA | 1336 |
| QY | 1021 | GAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGAACACTCTTCACATCTTGTGTTCTATA | 1080 |
| Db | 1337 | GAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGAACACTCTTCACATCTTGTGTTCTATA | 1396 |
| QY | 1081 | AACACTGAGAGAGACAGCAATGATTAATAAAAAACAGATGTTGGGAGAGAGACCACTGG | 1140 |
| Db | 1397 | AACACTGAGAGAGACAGCAATGATTAATAAAAAACAGATGTTGGGAGAGAGACCACTGG | 1456 |
| QY | 1141 | ATCGGCCCTCAGACTCAGAGCGTGAATAATGAATGGAAGTGGCTGATTTGGACATCTTCCA | 1200 |

Db 1457 ATGGGCTTCACAGACTGAGAGCGTGAATAAGAAAGTGGCGTGGATCTCCA 1516
 QY 1201 GACTACAAAATTTGGAAGCTGGACCGGATTAAGTGGTCAATGGCCATGGCCCGGA 1260
 Db 1517 GACTACAAAATTTGGAAGCTGGACCGGATTAAGTGGTCAATGGCCATGGCCCGGA 1576
 QY 1261 GAAGACTGTGCTGGTGTATTTATGCTGGCAGTGAACGATTTCCATGTGAAGAGCTC 1320
 Db 1577 GAAGACTGTGCTGGTGTATTTATGCTGGCAGTGAACGATTTCCATGTGAAGAGCTC 1636
 QY 1321 AATTAATCTATTGGCGAAAAAGACAGGAGACAGTACTGATCTGCATTA 1371
 Db 1637 AATTAATCTATTGGCGAAAAAGACAGGAGACAGTACTGATCTGCATTA 1687
 RESULT 5
 AAC66903
 ID AAC66903 standard; cDNA; 2641 BP.
 AC AAC66903;
 XX
 DT 27-MAR-2001 (first entry)
 DE Human EXMAD-14 coding sequence SEQ ID NO: 39.
 XX
 XX Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
 KM inflammation; reproductive disorder; cardiovascular disorder;
 KM immune disorder; musculoskeletal disorder; developmental disorder;
 KM gastrointestinal disorder; cell proliferation disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO200068380-A2.
 XX
 PD 16-NOV-2000.
 XX
 PF 10-MAY-2000; 2000MO-US12811.
 XX
 PR 11-MAY-1999; 99US-0133643.
 PR 23-AUG-1999; 99US-0150409.
 XX
 XX (INCYTE-) INCYTE GENOMICS INC.
 PA Bandman O, Hillman JL, Tang YF, Lai P, Yue H, Baughn MR, Lu DM;
 PI Azimzai Y;
 XX
 DR WPI: 2001-007395/01.
 DR P-PsDB; AAB27236.
 PT Isolated polynucleotide encoding extracellular matrix or
 PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
 PT preventing disorders associated with expression of EXMAD such as
 PT proliferative, immune and genetic disorders -
 XX
 PS Claim 4; Page 121-122; 129pp; English.
 XX
 XX The present invention provides the protein and coding sequences for 25
 CC novel extracellular matrix and adhesion-associated proteins (EXMADs).
 CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
 CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
 CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
 CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
 CC useful in the prevention and treatment of cancers, cell proliferation,
 CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
 CC gastrointestinal disorders and inflammation.
 XX
 SQ Sequence 2641 BP; 823 A; 628 C; 608 G; 582 T; 0 other;

QY 1 ATGAACAGCCAGCTCAACTCATTACAGAGTCAAGTGAAGAAATCAACCACTATCTCTCAA 60
 Db 567 ATGAACAGCCAGCTCAACTCATTACAGAGTCAAGTGAAGAAATCAACCACTATCTCTCAA 626
 QY 61 GCCAAGAGAGAGAAACCTGGAAGACCTTCAGAGACTTACACAAATATGAGAAATTAACA 120
 Db 627 GCCAAGAGAGAGAAACCTGGAAGACCTTCAGAGACTTACACAAATATGAGAAATTAACA 686
 QY 121 GCCATCAAGTTCAACCAACTGAGAGAAAGCTTCACAGCTCTTTGAGAGAGATATTTGAAC 180
 Db 687 GCCATCAAGTTCAACCAACTGAGAGAAAGCTTCACAGCTCTTTGAGAGAGATATTTGAAC 746
 QY 181 ATCAATAGCAATATACGTTTACACAGCCACACCTCGGAGAGTGAACAGCAATCTAAT 240
 Db 747 ATCAATAGCAATATACGTTTACACAGCCACACCTCGGAGAGTGAACAGCAATCTAAT 806
 QY 241 GAAGTCAAGAGCACTTGCACAGATACCCCTTACCAAAACACAGATATGATGACCTCTTG 300
 Db 807 GAAGTCAAGAGCACTTGCACAGATACCCCTTACCAAAACACAGATATGATGACCTCTTG 866
 QY 301 AATAATACCTTGAGCCCAACATCCGTTTGATTTCTCTCAGATGCAACAGATTTG 360
 Db 867 AATAATACCTTGAGCCCAACATCCGTTTGATTTCTCTCAGATGCAACAGATTTG 926
 QY 361 ATGAGGTGAGGTTAGACACTGAAGTACCACTTATCACTGATATGAAGAATGAAG 420
 Db 927 ATGAGGTGAGGTTAGACACTGAAGTACCACTTATCACTGATATGAAGAATGAAG 986
 QY 421 CTGATGACCTCCAAAGCATGCTGATCAATCAAAATTTTCAATACAGAGTCCACCG 480
 Db 987 CTGATGACCTCCAAAGCATGCTGATCAATCAAAATTTTCAATACAGAGTCCACCG 1046
 QY 481 GGCCCCAGAGGGTCCAAAGAGTGAAGAGATCCCAAGAGACCCCTGAGCCCAACTGAGCAAC 540
 Db 1047 GGCCCCAGAGGGTCCAAAGAGTGAAGAGATCCCAAGAGACCCCTGAGCCCAACTGAGCAAC 1106
 QY 541 AAGGACAGAAAGAGAGAAAGGGGAGCTGACACACTGCGCTCGGGTGAAGAGGC 600
 Db 1107 AAGGACAGAAAGAGAGAAAGGGGAGCTGACACACTGCGCTCGGGTGAAGAGGC 1166
 QY 601 CCAATTGACACAGCTGGTCCCGGAGAGCTGGGGGCAAGAGATCTAAAGGCTCCGAC 660
 Db 1167 CCAATTGACACAGCTGGTCCCGGAGAGCTGGGGGCAAGAGATCTAAAGGCTCCGAC 1226
 QY 661 GGCCCCAAAGGCTCCGTTGTTCCCTGGGAAAGCCCGGCTTCAAGGAGCCCACTGGAGAG 720
 Db 1227 GGCCCCAAAGGCTCCGTTGTTCCCTGGGAAAGCCCGGCTTCAAGGAGCCCACTGGAGAG 1286
 QY 721 CCAAGGCCCCCGGGGCGCCACAGGCAAGAGAGACTCCCGGCTCAAGGGCTCTCTGGC 780
 Db 1287 CCAAGGCCCCCGGGGCGCCACAGGCAAGAGAGACTCCCGGCTCAAGGGCTCTCTGGC 1346
 QY 781 TTCAGGAGGACTTTCAGGAGCACTGTTGGGAGCTGAGGAGCTCGAGCTCGAGGAGTGC 840
 Db 1347 TTCAGGAGGACTTTCAGGAGCACTGTTGGGAGCTGAGGAGCTCGAGCTCGAGGAGTGC 1406
 QY 841 GGCTTGCTTGGGGTATCCAGGCAATGCCAGGCCCCCAAGGGCCCCCGGCTCTCTGGCCCA 900
 Db 1407 GGCTTGCTTGGGGTATCCAGGCAATGCCAGGCCCCCAAGGGCCCCCGGCTCTCTGGCCCA 1466
 QY 901 TCAGGAGAGGTTGCTGCCCCCTGCGCTGCAAGATGAGGCAACCCCGGACCCGAGAGCAAT 960
 Db 1467 TCAGGAGAGGTTGCTGCCCCCTGCGCTGCAAGATGAGGCAACCCCGGACCCGAGAGCAAT 1526
 QY 961 GGCTGCGCCCGCTCACTGGAAGAACTTCAACAGCAAAATGCTACTATTTTCAAGTTGAGAA 1020
 Db 1527 AGCTGCGCCCGCTCACTGGAAGAACTTCAACAGCAAAATGCTACTATTTTCAAGTTGAGAA 1586
 QY 1021 GAAATTTTGAAGATGCAAAAGCTTTTCTGGAAGACAGCTTCACATCTTGTGTTTCATA 1080
 Db 1587 GAAATTTTGAAGATGCAAAAGCTTTTCTGGAAGACAGCTTCACATCTTGTGTTTCATA 1646
 QY 1081 AACACTAGAGAGAGAACAGCAATGATTAATAAAACAGATGTTAGGAGAGAGAGCCACTGG 1140

Query Match 99.9%; Score 1369.4; DB 22; Length 2641;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1647 AACACTAGAGAGACAGCAATGATATAAAAAACAGANTGTTAGGAGAGAGACCACTGG 1706
 QY 1141 ATCGGCTTCACAGACGTACAGAGGTGAATGAATGAAATGCTGGATGGACATCTCCA 1200
 Db 1707 ATCGGCTTCACAGACGTACAGAGGTGAATGAATGAAATGCTGGATGGACATCTCCA 1766
 QY 1201 GACTACAAAAATTTGGAAAGCTGACAGCCGGATTAATGCTGGATGGACATCTCCA 1260
 Db 1767 GACTACAAAAATTTGGAAAGCTGACAGCCGGATTAATGCTGGATGGACATCTCCA 1826
 QY 1261 GAGAGCTGTGCTGGTGTATTTATGCTGGGACATGAGACGATTTCCATGTGAAGACGTC 1320
 Db 1827 GAGAGCTGTGCTGGTGTATTTATGCTGGGACATGAGACGATTTCCATGTGAAGACGTC 1886
 QY 1321 AATACTTCATTTGGAAAAAGACAGGAGACAGTACTGCTCATCTCCATTA 1371
 Db 1887 AATACTTCATTTGGAAAAAGACAGGAGACAGTACTGCTCATCTCCATTA 1937

RESULT 6

ABO92072 ID ABO92072 standard; cDNA; 2929 BP.

AC ABO92072;

DT 04-OCT-2002 (first entry)

DE Human polynucleotide SEQ ID NO 69.

XX Human; cytostatic; antirheumatic; antiarthritic; vulnary; analgesic;
 KM antinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
 KM neuroprotective; nootropic; osteopathic; haemostatic; vasotonic;
 KM antituber; fungicide; antidiabetic; antistatic; antiallergic;
 KM immunostimulant; antiparasitic; secreted protein; transmembrane protein;
 KM cytokine; cell proliferation; cell differentiation; neuroimmune disease;
 KM stem cell; growth factor; nervous system disease; neuropathy;
 KM Alzheimer's disease; Parkinson's disease; Huntington's disease;
 KM osteoporosis; severe combined immunodeficiency; SCID; infection;
 KM multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.

XX Homo sapiens.

OS US2002065394-A1.

PN 30-MAY-2002.

XX 22-DEC-2000; 2000US-0745763.

PR 18-MAR-1998; 98US-0040963.

XX (JACO/) JACOBS K.

PA (MCCO/) MCCOY J M.

PA (LAVA/) LAVALLIE E R.

PA (COLL/) COLLINS-RACIE L A.

PA (EVAN/) EVANS C.

PA (MERB/) MERBERG D.

PA (TREA/) TREACY M.

PA (SPAU/) SPAULDING V.

PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M, Spaulding V;

DR WPI: 2002-582343/62.

DR P-PSDB; ABB61859.

XX Novel secreted or transmembrane protein and polynucleotide encoding the

XX protein, useful for diagnosis and treatment of neurological disorders,

XX cancer, autoimmune diseases, bone disorders and lung or liver fibrosis

XX Claim 250; Page 231-232; 284pp; English.

CC The invention relates to human secreted or transmembrane protein (I),
 CC their fragments and is encoded by specific complementary deoxyribonucleic
 CC acid (cDNA) inserts (II), where the protein is substantially free from
 CC other mammalian proteins. (I) are useful for preventing, treating or
 CC ameliorating a medical condition, especially immunological treatment or
 CC prevention of tumors. (I) exhibits activity relating to angiogenesis,
 CC cytokine, cell proliferation, cell differentiation, antinflammatory,
 CC stem cell growth factor activity and activin or inhibin-related
 CC activities. (I) can be used to manipulate stem cells in culture to give
 CC rise to neuroepithelial cells that can be used to augment or replace
 CC cells damaged by illness, autoimmune disease, accidental damage or
 CC genetic disorders. (I) induces the proliferation of neural cells and
 CC regeneration of nerve and brain tissue and is useful for the treatment of
 CC central and peripheral nervous system diseases and neuropathies, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic
 CC activity, regulation of haematopoiesis and is useful for treating myeloid
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
 CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,
 CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
 CC periodontal disease. (I) is also useful for gut protection or
 CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
 CC in various tissues, various immune deficiencies and disorders including
 CC severe combined immunodeficiency (SCID), bacterial or fungal infections,
 CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
 CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
 CC such as asthma or other respiratory problems. (II) is useful to express
 CC recombinant protein, as markers for tissues in which the corresponding
 CC protein is preferentially expressed and in gene therapy. The present
 CC sequence is that of a polynucleotide of the invention.

Sequence 2929 BP; 901 A; 699 C; 692 G; 637 T; 0 other;

Query Match 99.9%; Score 1369.4; DB 24; Length 2929;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAACAGCAGGTCAATTCATACAGGTGAGTGAAGAACATCCACTATCTCTCAA 60
 Db 921 ATGAACAGCAGGTCAATTCATACAGGTGAGTGAAGAACATCCACTATCTCTCAA 980
 QY 61 GCCAAGAGCAGACACTGAAAGACCTGAGGACTTACACAAAGATGACAGATAGACA 120
 Db 981 GCCAAGAGCAGACACTGAAAGACCTGAGGACTTACACAAAGATGACAGATAGACA 1040
 QY 121 GCCATCAAGTCAACCACTGAGGAGACCTCTTCCAGCGATATTGTGAAC 180
 Db 1041 GCCATCAAGTCAACCACTGAGGAGACCTCTTCCAGCGATATTGTGAAC 1100
 QY 181 ATCATAGCAATATGATACACAGCCACACCTGCGAGCGTGCACGATATTAAT 240
 Db 1101 ATCATAGCAATATGATACACAGCCACACCTGCGAGCGTGCACGATATTAAT 1160
 QY 241 GAAGTGAAGACCACTGACAGATATCTTACCAACACACAGATATCTGACCTCTTG 300
 Db 1161 GAAGTGAAGACCACTGACAGATATCTTACCAACACACAGATATCTGACCTCTTG 1220
 QY 301 AATATATCCCTGGCCACATCCGTTGGATTCTGTTCTCAGATGCAACAAGATTG 360
 Db 1221 AATATATCCCTGGCCACATCCGTTGGATTCTGTTCTCAGATGCAACAAGATTG 1280
 QY 361 ATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 420
 Db 1281 ATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1340
 QY 421 CTAGTAGTCCACAGATGTCAGTCAATCAAGATTTTACATACTACAGGTCACCG 480
 Db 1341 CTAGTAGTCCACAGATGTCAGTCAATCAAGATTTTACATACTACAGGTCACCG 1400
 QY 481 GGCCCCAGGGGTCACAGAGGTGACAGAGATCCACAGGACCCCTGCGCCAACTGGCAAC 540
 Db 1401 GGCCCCAGGGGTCACAGAGGTGACAGAGATCCACAGGACCCCTGCGCCAACTGGCAAC 1460

| | | | |
|----------|-------------|---|------|
| QY | 541 | AAGGACAGAAAGAGAGAGAGGGGGAGCCTGACACCACTGGCCCTGGCGGTGAGAGAGGC | 600 |
| Db | 1461 | AAGGACAGAAAGAGAGAGAGGGGGAGCCTGGACCACTGGCCCTGGCGGTGAGAGAGGC | 1520 |
| QY | 601 | CCAAATTGAGACAGCTGGTCCCTCCGGAGAGCGTGGCGCAAGAGATCTAAAGAGCTCCAG | 660 |
| Db | 1521 | CCAAATTGAGACAGCTGGTCCCTCCGGAGAGCGTGGCGCAAGAGATCTAAAGAGCTCCAG | 1580 |
| QY | 661 | GGCCCCAAAGGCTCCGTGGTTCCCTGGGAAGCCGGGCCCTCAGGGCCCCAGTGGGGAC | 720 |
| Db | 1581 | GGCCCCAAAGGCTCCGTGGTTCCCTGGGAAGCCGGGCCCTCAGGGCCCCAGTGGGGAC | 1640 |
| QY | 721 | CCAGAGCCCCCGGGGCCACACAGGCAAAAGAGGGGACATCCCGGGCCCTCAGGGCCCTCGGGC | 780 |
| Db | 1641 | CCAGAGCCCCCGGGGCCACACAGGCAAAAGAGGGGACATCCCGGGCCCTCAGGGCCCTCGGGC | 1700 |
| QY | 781 | TTTCAGGAGACTTCAGGGCACCCTGTGGGGAGCCTGGGGGTGCTCGACCTCGGGGACTGCCA | 840 |
| Db | 1701 | TTTCAGGAGACTTCAGGGCACCCTGTGGGGAGCCTGGGGGTGCTCGACCTCGGGGACTGCCA | 1760 |
| QY | 841 | GGCTTGGCTGGGGTACCGGACATGCCAGGGCCCCCAAGGGCCCCCGGGCCCTCCTGGGCCA | 900 |
| Db | 1761 | GGCTTGGCTGGGGTACCGGACATGCCAGGGCCCCCAAGGGCCCCCGGGCCCTCCTGGGCCA | 1820 |
| QY | 901 | TCAGAGACCGGTGGTCCCGGCGCTTCGAGAAATGAGACCACCCGGGACCGGAGAGACAAT | 960 |
| Db | 1821 | TCAGAGACCGGTGGTCCCGGCGCTTCGAGAAATGAGACCACCCGGGACCGGAGAGACAAT | 1880 |
| QY | 961 | GGCTGCCCGGCTCACTGGAGAACTTCACAGACCAATGCTACTATTTTTCAGTTGAGAAA | 1020 |
| Db | 1881 | AGCTGCCCGGCTCACTGGAGAACTTCACAGACCAATGCTACTATTTTTCAGTTGAGAAA | 1940 |
| QY | 1021 | GAATTTTGTAGAGATGCAAGGCTTTCGTGGAAGACAGTCTTCACATCTGTTTGATA | 1080 |
| Db | 1941 | GAATTTTGTAGAGATGCAAGGCTTTCGTGGAAGACAGTCTTCACATCTGTTTGATA | 2000 |
| QY | 1081 | AACACTGAGAGAGAACAGCAATGGATMAAAACACAGATGGTAGGAGAGAGAGCACTGG | 1140 |
| Db | 2001 | AACACTGAGAGAGAACAGCAATGGATMAAAACACAGATGGTAGGAGAGAGAGCACTGG | 2060 |
| QY | 1141 | ATGGGCGCTCACAGACTCAGAGCGCTGAAATTAATGGAAGTGGCTGGATGGACATCTGCCA | 1200 |
| Db | 2061 | ATGGGCGCTCACAGACTCAGAGCGCTGAAATTAATGGAAGTGGCTGGATGGAGCATCTGCCA | 2120 |
| QY | 1201 | GACTACAAAAAATTTGGAAGCGTGGAGCCGGATTAATCTGGGGGTCATGGGCATGGGCCAGGA | 1260 |
| Db | 2121 | GACTACAAAAAATTTGGAAGCGTGGAGCCGGATTAATCTGGGGGTCATGGGCATGGGCCAGGA | 2180 |
| QY | 1261 | GAAAGCTGTGTGGGTGATTATATGCTGGGCGAGTGGAAACGATTTCCAAATGTGAAGACGTC | 1320 |
| Db | 2181 | GAAAGCTGTGTGGGTGATTATATGCTGGGCGAGTGGAAACGATTTCCAAATGTGAAGACGTC | 2240 |
| QY | 1321 | AATAACTCTCAATTTGCGGAAAAAGACAGGAGAGACAGTACGTCTATCTGCGATT | 1371 |
| Db | 2241 | AATAACTCTCAATTTGCGGAAAAAGACAGGAGAGACAGTACGTCTATCTGCGATT | 2291 |
| RESULT 7 | | | |
| ID | AAV55746 | standard; cDNA; 2930 BP. | |
| XX | AAV55746; | | |
| XX | 23-MAR-1999 | (first entry) | |
| XX | | Human secreted protein clone bv227_1 coding sequence. | |
| KM | | Secreted protein; human; nutritional supplements; immune stimulant; | |
| KM | | immune suppressor; hematopoietic regulator; tissue growth; chemotaxis | |
| KM | | activin/inhibin; chemokinesis; haemostasis; thrombolysis; | |
| KM | | receptor/19and activity; anti-inflammatory; tumour inhibitor; | |
| KM | | cadherin/tumour invasion suppressor; ds. | |

| | |
|----------------------------|---|
| xx | Homo sapiens. |
| os | |
| xx | |
| EH | Location/Qualifiers |
| FT | CDS 67..693 |
| ET | /*tag= a |
| PN | |
| xx | W09855614-A2. |
| xx | |
| PD | 10-DEC-1998. |
| XX | |
| PF | 01-JUN-1998; 98WO-US11210. |
| XX | |
| PR | 29-MAY-1998; 98US-0087255. |
| PR | 04-JUN-1997; 97US-0868696. |
| PR | 04-JUN-1997; 97US-0868697. |
| PR | 04-JUN-1997; 97US-0868698. |
| PR | 04-JUN-1997; 97US-0868698. |
| PR | 04-JUN-1997; 97US-0868899. |
| PR | 04-JUN-1997; 97US-0868900. |
| PR | 04-JUN-1997; 97US-0869191. |
| PR | 04-JUN-1997; 97US-0869192. |
| PR | 04-JUN-1997; 97US-0869193. |
| PR | 04-JUN-1997; 97US-0869194. |
| PA | (GENY) GENETICS INST INC. |
| XX | |
| PI | Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER; |
| PI | Mccoy JM, Racie LA, Spaulding V, Treacy M; |
| DR | MP1. 1999-059912/05. |
| DR | P-PDB; AAM73628. |
| xx | |
| PT | New polynucleotides encoding secreted human proteins - derived from |
| PT | human foetal brain, adult brain, foetal kidney, adult ovary, adult |
| PT | retina, adult placenta or adult uterus cDNA libraries |
| PS | Claim 26; Page 87-88; 127pp; English. |
| CC | This sequence encodes a human secreted protein of the invention. |
| CC | This DNA sequence was isolated from a human adult brain cDNA |
| CC | library, and was designated clone bv227.1. The DNAs and proteins |
| CC | are predicted to have biological activities which would make them |
| CC | suitable for treating, preventing or ameliorating medical conditions in |
| CC | humans and animals, although no supporting data is given. Suggested |
| CC | activities include nutritional sources or supplements, immune |
| CC | stimulating or suppressing activity, haematopoiesis regulating activity, |
| CC | tissue growth activity, activin/inhibin activity, |
| CC | chemoattractant/chemokinetic activity, haemostatic and thrombolytic activity, |
| CC | receptor/ligand activity, anti-inflammatory activity, cadherin/tumour |
| CC | invasion suppressor activity, and tumour inhibition activity. The DNAs |
| CC | are also stated to be useful for gene therapy. A host cell transfected |
| CC | with the DNA, or its subfragments and variants is useful for recombinant |
| CC | production of the human secreted protein clones. |
| XX | |
| SQ | Sequence 2930 BP; 902 A; 699 C; 692 G; 637 T; 0 other: |
| Query Match | 99.9%; Score 1369.4; DB 20; Length 2930; |
| Best Local Similarity | 99.9%; Pred. No. 0; Matches 1; Indels 0; Gaps 0; |
| Matches 1370; Conservative | 0; Mismatches |
| OY | 1 ATGACAGCGACCTCAACTCATTTCACAGTGCAGATGAGAACATCATCACCTATCTCTCAA 60 |
| Db | 921 ATGAACAGCGACCTCAACTCATTTCACAGTGCAGATGAGAACATCATCACCTATCTCTCAA 980 |
| OY | 61 GCCAAGAGACAGACCTGAAAGACCTGCAGACTTACACAAAGATGACAGAGATGAGACA 120 |
| Db | 981 GCCAAGAGAGACAAACCTGAAAGACCTGCAGACTTACACAAAGATGACAGAGATGAGACA 1040 |
| OY | 121 GCATCAGATTCACCAACTGGAGAAGCGTTCCAGCTCTTGAGACGATATTGTGAAAC 180 |
| Db | 1041 GCATCAGATTCACCAACTGGAGAAGCGTTCCAGCTCTTGAGACGATATTGTGAAAC 1100 |

QY 181 ATCATTTAGCAATATTCAGTTTACACAGCCACACCTGCGAGCGCTGACACCAATCTAAAT 240
 Db 1101 ATCATTTAGCAATATTCAGTTTACACAGCCACACCTGCGAGCGCTGACACCAATCTAAAT 1160
 QY 241 GAAGTGAAGACCACTTGCACAGATACCCTTACCACACAGATGATGCTGACCTCTTG 300
 Db 1161 GAAGTGAAGACCACTTGCACAGATACCCTTACCACACAGATGATGCTGACCTCTTG 1220
 QY 301 AATATATACCTGCGCCAAACATCCGTTGGATTTCTTCTCTCGAGATGCAACAAGATTG 360
 Db 1221 AATATATACCTGCGCCAAACATCCGTTGGATTTCTTCTCTCGAGATGCAACAAGATTG 1280
 QY 361 ATGAGGTCGAGGTTTACACTGAAGTACCACTTATCAGTGTATTTGAAGAAATGAAG 420
 Db 1281 ATGAGGTCGAGGTTTACACTGAAGTACCACTTATCAGTGTATTTGAAGAAATGAAG 1340
 QY 421 CTAGTGAAGTCCACAGCATGCTGACCTCATCAGAAATTTTACAAATCTACAGGTCACCG 480
 Db 1341 CTAGTGAAGTCCACAGCATGCTGACCTCATCAGAAATTTTACAAATCTACAGGTCACCG 1400
 QY 481 GGGCCCCAGGGGTCACAGAGGTGACAGAGATCCAGGAGCCCTGCGCCCACTGGCAAC 540
 Db 1401 GGGCCCCAGGGGTCACAGAGGTGACAGAGATCCAGGAGCCCTGCGCCCACTGGCAAC 1460
 QY 541 AAGGACAGAAAAGAGAGAGGAGGAGCCTGACACACTGGCCCTGCGGTGAGAGAGGC 600
 Db 1461 AAGGACAGAAAAGAGAGAGGAGGAGCCTGACACACTGGCCCTGCGGTGAGAGAGGC 1520
 QY 601 CCAATTTGACACAGCTGCTCCCCCGAGAGAGCTGCGGCAAGATCTAAAGGCTCCAG 660
 Db 1521 CCAATTTGACACAGCTGCTCCCCCGAGAGAGCTGCGGCAAGATCTAAAGGCTCCAG 1580
 QY 661 GGGCCCCAAGGGTCCCGTGGTTCCCTGGGAAAGCCCGCCCTGAGGGCCCACTGGGGAG 720
 Db 1581 GGGCCCCAAGGGTCCCGTGGTTCCCTGGGAAAGCCCGCCCTGAGGGCCCACTGGGGAG 1640
 QY 721 CCAGGCCCCCGGGCCACACAGGCAAGAGAGACTCCCGGCCCTGAGGGCCCTGCTGGC 780
 Db 1641 CCAGGCCCCCGGGCCACACAGGCAAGAGAGACTCCCGGCCCTGAGGGCCCTGCTGGC 1700
 QY 781 TTCCAGGACCTTCAGGGGACCCGTTGGGAGCCTGGGGTGCCTGGACCTGGGGAGCTGCCA 840
 Db 1701 TTCCAGGACCTTCAGGGGACCCGTTGGGAGCCTGGGGTGCCTGGACCTGGGGAGCTGCCA 1760
 QY 841 GGCTTGGCCGCGGGTACCAAGCATGCCAGGCCCCCAAGGGCCCTGCTGGGCCA 900
 Db 1761 GGCTTGGCCGCGGGTACCAAGCATGCCAGGCCCCCAAGGGCCCTGCTGGGCCA 1820
 QY 901 TCAGAGAGCGGTGTCGCCCTGGCCCTGAGAAATGAGCCAAACCCCGGACCGAGAGCAAT 960
 Db 1821 TCAGAGAGCGGTGTCGCCCTGGCCCTGAGAAATGAGCCAAACCCCGGACCGAGAGCAAT 1880
 QY 961 GGCTGCCCCCTCTACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1020
 Db 1881 GGCTGCCCCCTCTACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1940
 QY 1021 GAAATTTTGGAGGATGCAAGCTTTCTGTGAAGACAAGCTTCACATCTGTTTTCATA 1080
 Db 1941 GAAATTTTGGAGGATGCAAGCTTTCTGTGAAGACAAGCTTCACATCTGTTTTCATA 2000
 QY 1081 AACACTAGAGAGAACAGCAATGATTAATAAACAAGATGATGAGAGAGAGAGAGCACTGG 1140
 Db 2001 AACACTAGAGAGAACAGCAATGATTAATAAACAAGATGATGAGAGAGAGAGAGCACTGG 2060
 QY 1141 ATCGGCTCAGACACTCAGAGCTGAAATGAATGAAGTGGCTGATGGAGCATCTCCA 1200
 Db 2061 ATCGGCTCAGACACTCAGAGCTGAAATGAATGAAGTGGCTGATGGAGCATCTCCA 2120
 QY 1201 GACTACAAAATTTGAAAAGCTGGACAGCCGGAATAACTGGGGTATGAGGATGGGCCAGGA 1260
 Db 2121 GACTACAAAATTTGAAAAGCTGGACAGCCGGAATAACTGGGGTATGAGGATGGGCCAGGA 2180
 QY 1261 GAAGACTGTGCTGGGTGATTATGCTGGGCAAGTGAAGCATTTCCAAATGTGAAGAGCTC 1320

Db 2181 GAAGACTGTGCTGGGTGATTATGCTGGGAGAGATGGAACATGTAAGAGCTC 2240
 QY 1321 AATTAATCTTCTTTCGCAAAAAGACAGGAGACAGTACTGATCATCTGATTA 1371
 Db 2241 AATTAATCTTCTTTCGCAAAAAGACAGGAGACAGTACTGATCATCTGATTA 2291
 RESULT 8
 ID AA158842 standard; cDNA; 2318 BP.
 AC
 XX
 AA158842;
 XX
 DT 22-OCT-2001 (first entry) -
 XX
 DE Human polynucleotide SEQ ID NO 1045.
 DE
 KW Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukemia; ss.
 KW
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 PD
 XX 26-JUL-2001.
 PF
 XX 26-DEC-2000; 2000WO-US34263.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Dirmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR P-PsDB; AAM39686.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PT
 XX
 PS
 XX
 Claim 1; SEQ ID NO 1045; 10078bp; English.
 CC
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AA442213) with nootropic,
 CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC actinin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC

Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other;

Query Match 99.1%; Score 1358.4; DB 22; Length 2318;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1370; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGACAGCCAGCTCACTATTACAGGTACATGAGAAATCACCATTCTCTCAA 60
DB 317 ATGAAAGCCAGCTCACTATTACAGGTACATGAGAAATCACCATTCTCTCAA 376
QY 61 -GCCAAGCAGAGAACTGAAAGACTGCAAGACTTACACAAAGATGAGAAATAGAC 119
DB 377 GGGCAAGCAGAGAACTGAAAGACTGCAAGACTTACACAAAGATGAGAAATAGAC 436
QY 120 AGCCATCAAGTTCAACCACTGAGGAACCTCTTTCAGAGCGATATTGTGAA 179
DB 437 AGCCATCAAGTTCAACCACTGAGGAACCTCTTTCAGAGCGATATTGTGAA 496
QY 180 CATCATTAAGCAATTCAGTTACACAGCCACACCTGCGAGCGTGAACCATCTTAA 239
DB 497 CATCATTAAGCAATTCAGTTACACAGCCACACCTGCGAGCGTGAACCATCTTAA 556
QY 240 TGAAGTCAGAGACCACTGACAGATACCTTACCAACACACAGATATCTGACCTCTT 299
DB 557 TGAAGTCAGAGACCACTGACAGATACCTTACCAACACACAGATATCTGACCTCTT 616
QY 300 GAATTAATACCTGCGCAACATCCGTTTGATCTGTTCTCTCAGATGCAACAAGATT 359
DB 617 GAATTAATACCTGCGCAACATCCGTTTGATCTGTTCTCTCAGATGCAACAAGATT 676
QY 360 GATGAGGTCCAGGTTAACACTGAAAGTACCACTTATCACTGATTTTGAAGAAATGAA 419
DB 677 GATGAGGTCCAGGTTAACACTGAAAGTACCACTTATCACTGATTTTGAAGAAATGAA 736
QY 420 GCTAGTACTCCAGAGATGTCAGTCAATCAAGATTTTACATTAATACAAAGTCCACC 479
DB 737 GCTAGTACTCCAGAGATGTCAGTCAATCAAGATTTTACATTAATACAAAGTCCACC 796
QY 480 GGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCCAAGGACCCCTGCGCCAACTGGCAA 539
DB 797 GGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCCAAGGACCCCTGCGCCAACTGGCAA 856
QY 540 CAAGGGACAAAGAGAGAGAGGGGAGCCTGGACCACTGCGCCCTGCGGTGAGAGAG 599
DB 857 CAAGGGACAAAGAGAGAGAGGGGAGCCTGGACCACTGCGCCCTGCGGTGAGAGAG 916
QY 600 CCCAATTGAGACAGCTGTCCTCCCGGAGAGCGTGGCGGCAAAAGATCTAAAGCTCCCA 659
DB 917 CCCAATTGAGACAGCTGTCCTCCCGGAGAGCGTGGCGGCAAAAGATCTAAAGCTCCCA 976
QY 660 GGGCCCCAAAGGCTCCGCTGCTCCCTGGGAAAGCCCGGCTTCAGGGCCCCAGTGGGGA 719
DB 977 GGGCCCCAAAGGCTCCGCTGCTCCCTGGGAAAGCCCGGCTTCAGGGCCCCAGTGGGGA 1036
QY 720 CCCAGGCCCCCGGGCCCCACCAAGGAAAGAGGACTCCCGGCCCCCTAGGGCCCCCTG 779
DB 1037 CCCAGGCCCCCGGGCCCCACCAAGGAAAGAGGACTCCCGGCCCCCTAGGGCCCCCTG 1096
QY 780 CTTTCAGAGGACTTCAGGGACCGCTTGGGAGACCTGGGGTGCCTGACCTCGGGGACTGCC 839
DB 1097 CTTTCAGAGGACTTCAGGGACCGCTTGGGAGACCTGGGGTGCCTGACCTCGGGGACTGCC 1156
QY 840 AGGCTTCCTGCGGGTACCAAGCATGCCAGGCCCAAGAGGCCCGCCCGCTCTCTGCGCC 899
DB 1157 AGGCTTCCTGCGGGTACCAAGCATGCCAGGCCCAAGAGGCCCGCCCGCTCTCTGCGCC 1216
QY 900 ATCAGAGAGCGGTGTCGCCCTGCGCGATGAGCCCAACCCCGGACCGGAGGACAA 959
DB 1217 ATCAGAGAGCGGTGTCGCCCTGCGCGATGAGCCCAACCCCGGACCGGAGGACAA 1276
QY 960 TGGCTGCCCGCTCACTGGAAGAACTCACAGCAAAATGCTACTATTCTTTCAGTTGAGAA 1019
DB 1277 TGGCTGCCCGCTCACTGGAAGAACTCACAGCAAAATGCTACTATTCTTTCAGTTGAGAA 1336

QY 1020 AGAATTTTGGAGATGCAAGACTTTTCTGTGAAGACAGTCTTCACATCTGTTTCAT 1079
DB 1337 AGAATTTTGGAGATGCAAGACTTTTCTGTGAAGACAGTCTTCACATCTGTTTCAT 1396
QY 1080 AAACACTAGAGAGAACAGCAATGATAAAAAACAATGTTAGGAGAGAGAGCCACTG 1139
DB 1397 AAACACTAGAGAGAACAGCAATGATAAAAAACAATGTTAGGAGAGAGAGCCACTG 1456
QY 1140 GATCGGCTCCACAGACTCAGAGCCTGAAATTAATGAAGTGGCTGATGGACATCTCC 1199
DB 1457 GATCGGCTCCACAGACTCAGAGCCTGAAATTAATGAAGTGGCTGATGGACATCTCC 1516
QY 1200 AGACTACAAAATTTGAAAGCTGACAGCCGGATTAATCTGGGCTCATGGCCATGGCCAGG 1259
DB 1517 AGACTACAAAATTTGAAAGCTGACAGCCGGATTAATCTGGGCTCATGGCCATGGCCAGG 1576
QY 1260 AGAAGCTGCTGCTGGTGAATTTGCTGGGCACTGGAACGATTTCCAAATGTGAAGCT 1319
DB 1577 AGAAGCTGCTGCTGGTGAATTTGCTGGGCACTGGAACGATTTCCAAATGTGAAGCT 1636
QY 1320 CAATTAATCTCAATTTGCGAAAAAGACAGAGAGACAGTACTGTCACTCATTTA 1371
DB 1637 CAATTAATCTCAATTTGCGAAAAAGACAGAGAGACAGTACTGTCACTCATTTA 1688

RESULT 9
ABL95574
ID ABL95574 standard; cDNA; 2005 BP.
XX
AC ABL95574;
XX
DT 19-JUL-2002 (first entry)
XX
DE Human angiogenesis related cDNA PRO7223 SPQ ID NO: 27.
XX
KW Human: angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
KW cardiant; cytoslastic; antiangiogenic; hypotensive; vulnerary;
KW antiarteriosclerotic; gene; ss.
XX
OS Homo sapiens.
PN WO200208284-A2.
XX
PD 31-JAN-2002.
XX
PF 09-JUL-2001; 2001MO-US21735.
XX
PR 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220624P.
PR 28-JUL-2000; 2000US-220664P.
PR 02-AUG-2000; 2000MO-US20710.
PR 17-AUG-2000; 2000US-222695P.
PR 23-AUG-2000; 2000US-0643657.
PR 24-AUG-2000; 2000MO-US2352.
PR 24-AUG-2000; 2000MO-US23328.
PR 07-SEP-2000; 2000US-230978P.
PR 15-SEP-2000; 2000US-000000P.
PR 18-SEP-2000; 2000US-0664610.
PR 18-SEP-2000; 2000US-0665350.
PR 24-OCT-2000; 2000US-242922P.
PR 08-NOV-2000; 2000US-0709238.
PR 08-NOV-2000; 2000MO-US30952.
PR 10-NOV-2000; 2000MO-US30873.
PR 01-DEC-2000; 2000MO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000MO-US34956.
PR 22-JAN-2001; 2001US-0767609.
PR 28-FEB-2001; 2001US-0796498.
PR 28-FEB-2001; 2001MO-US06520.
PR 01-MAR-2001; 2001MO-US06666.
PR 09-MAR-2001; 2001US-0802706.

QY 1321 AATACTTATTTTGCAGAAAAAGACAGGACAGTACTGTATCTGCATTAA 1371
DB 1453 AATACTTATTTTGCAGAAAAAGACAGGACAGTACTGTATCTGCATTAA 1503
RESULT 10
ID ABL88085
ID ABL88085 standard; cDNA; 2005 BP.
XX ABL88085;
AC
XX
DT 16-MAY-2002 (first entry)
XX
DE Human PRO7223 cDNA sequence SEQ ID NO:27.
XX
KW Human: angiogenesis; cardiact; cyostatic; antiangiogenic; hypotensive;
KW vulnerary; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;
KW angiotensin disorder; cardiac hypertrophy; atherosclerosis; hypertension;
KW age-related macular degeneration; arterial stenosis; angina;
KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;
KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
KW wound healing; chromosome mapping; gene mapping; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200200690-A2.
XX
PD 03-JAN-2002.
XX
PF 20-JUN-2001; 2001WO-US19692.
XX
XX 23-JUN-2000; 2000US-213637P.
PR 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220644P.
PR 02-AUG-2000; 2000WO-US20710.
PR 17-AUG-2000; 2000US-0643657.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 07-SEP-2000; 2000US-230978P.
PR 18-SEP-2000; 2000US-0664610.
PR 18-SEP-2000; 2000US-0665350.
PR 24-OCT-2000; 2000US-242922P.
PR 08-NOV-2000; 2000US-0709238.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32878.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 22-JAN-2001; 2001US-0767609.
PR 28-FEB-2001; 2001US-0796498.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808889.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 25-MAY-2001; 2001WO-US17092.
PR 30-MAY-2001; 2001US-0870574.
PR 30-MAY-2001; 2001WO-US17443.
PR 01-JUN-2001; 2001WO-US17800.
XX
XX (GETH) GENENTECH INC.
PI Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;

PI Stephan JF, Watanabe CK, Williams PM, Wood WT, Ye W;
XX WPI: 2002-090516/12.
DR P-PSDB; ABB84830.
XX
PT One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or angiogenic disorders in a mammal -
XX
PS Claim 2; Fig 27; 565pp; English.
XX
CC ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to
CC ABB85003. The PRO proteins and polynucleotides have cardiact, cyostatic,
CC antiangiogenic, hypotensive, vulnerary and antiarteriosclerotic
CC activities, and can be used in gene therapy. The PRO polynucleotides,
CC proteins, agonists and antagonists are useful for treating or diagnosing
CC a cardiovascular, endothelial or angiogenic disorder in a mammal,
CC e.g. cardiac hypertrophy, trauma, cancer, age-related macular
CC degeneration, atherosclerosis, hypertension, arterial stenosis,
CC rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis,
CC lymphangitis, tumour angiogenesis (such as breast carcinoma and liver
CC carcinoma) and wound healing. The PRO polynucleotides have applications
CC in molecular biology, including use as hybridisation probes, and in
CC chromosome and gene mapping. ABL88259 to ABL88267 represent primers and
CC probes used in the exemplification of the present invention.
XX
SQ Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;
XX
Query Match 99.0%; Score 1357.4; DB 24; Length 2005;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1369; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 ATGACAGCAGCTCACTCACTATTCACAGGACAGTGAAGACATCACTATCTCTCAA 60
DB 134 ATGACAGCAGCTCACTCACTATTCACAGGACAGTGAAGACATCACTATCTCTCAA 193
QY 61 GCCACGAGCAGACCTGAAGACCTGACAGACTTACACAAAGATCCAGAGATAGACA 120
DB 194 GCCACGAGCAGACCTGAAGACCTGACAGACTTACACAAAGATCCAGAGATAGACA 253
QY 121 GCCATCAAGTTCAACCACTGGAGAGAGCTTCCACTCTTTGAGCGGATTTGGAAC 180
DB 254 GCCATCAAGTTCAACCACTGGAGAGAGCTTCCACTCTTTGAGCGGATTTGGAAC 313
QY 181 ATCATTAGCAATATCACTTACACAGCCACACCTCGGAGCGGTGACCAATCTTAAT 240
DB 314 ATCATTAGCAATATCACTTACACAGCCACACCTCGGAGCGGTGACCAATCTTAAT 373
QY 241 GAAGTCAGGACCACTTGACAGATACCTTTACCAACACAGATATGTGACCTCTTG 300
DB 374 GAAGTCAGGACCACTTGACAGATACCTTTACCAACACAGATATGTGACCTCTTG 432
QY 301 AATAATACCTGGCCAAACATCCGTTGGATGCTGTTCTCTCGATGACCAAGATTGG 360
DB 433 AATAATACCTGGCCAAACATCCGTTGGATGCTGTTCTCTCGATGACCAAGATTGG 492
QY 361 ATGAGGTCGAGGTTAGACACTGAGTACCACTTATCACTGATTTGGAAGAAATGAAG 420
DB 493 ATGAGGTCGAGGTTAGACACTGAGTACCACTTATCACTGATTTGGAAGAAATGAAG 552
QY 421 CTAGTAGACTCAAGCATGTGTCAGCTCATCAGAAATTTTCAATTAATCTACAGGTCACCG 480
DB 553 CTAGTAGACTCAAGCATGTGTCAGCTCATCAGAAATTTTCAATTAATCTACAGGTCACCG 612
QY 481 GGCCCGAGGGGTCAGAGGTGACAGAGATCCAGAGGACCCCTGCGCAACTGGCCAGC 540
DB 613 GGCCCGAGGGGTCAGAGGTGACAGAGATCCAGAGGACCCCTGCGCAACTGGCCAGC 672
QY 541 AAGGACAGAAAGAGAGAGAGGGGAGCCTGAGCACTTGCCCTGCGGAGAGAGGC 600
DB 673 AAGGACAGAAAGAGAGAGAGGGGAGCCTGAGCACTTGCCCTGCGGAGAGAGGC 732
QY 601 CCAATTGACCACTGTGTCGCCCGGAGAGAGCGTGGCGCAAGATCTAAAGGCTCCAG 660

| | | | |
|----|------|--|------|
| Db | 472 | GGCCCTCTGGCTTCACAGGACTTTCAGGGGACCGTTGGGAGCCCTGGGTGCTGGACT | 531 |
| QY | 829 | CGGGGACCTGCCAAGGCTTGCTGGGGTACCAAGCATGCCAGGCCCCCAAGGCCGCCCGGC | 888 |
| Db | 532 | CGGGGACCTGCCAAGGCTTGCTGGGGTACCAAGCATGCCAGGCCCCCAAGGCCGCCCGGC | 591 |
| QY | 889 | CCCTCTGGCCATCAGAGAGGATGGTGGCCCTGGGCCCTCACAATGAGGCACGCCGGCA | 948 |
| Db | 592 | CTCTCTGGCCATCAGAGAGGATGGTGGCCCTGGGCCCTCACAATGAGGCACGCCGGCA | 651 |
| QY | 949 | CCGGAGACAAATGGCTGCGCCCTCACTGGAGAACTTTCACAGACAAATGCTACTATTTT | 1008 |
| Db | 652 | CCGGAGACAAATGGCTGCGCCCTCACTGGAGAACTTTCACAGACAAATGCTACTATTTT | 711 |
| QY | 1009 | TCAGTTGAGAAAGAAATTTTGTGAGATGCAAGACTTTCTGTGAAGACAAAGTCTTCACAT | 1068 |
| Db | 712 | TCAGTTGAGAAAGAAATTTTGTGAGATGCAAGACTTTCTGTGAAGACAAAGTCTTCACAT | 771 |
| QY | 1069 | CTTGTTTTTCAATAACCTAGAGAGGAACAGCAATGGATATAAAAACAGATGGTAGGGAGA | 1128 |
| Db | 772 | CTTGTTTTTCAATAACCTAGAGAGGAACAGCAATGGATATAAAAACAGATGGTAGGGAGA | 831 |
| QY | 1129 | GAGAGCCACATGATGGCCCTCACAAGACTAGAGCGTGAAGATGAATGGAGTGGCTGGAT | 1188 |
| Db | 832 | GAGAGCCACATGATGGCCCTCACAAGACTAGAGCGTGAAGATGAATGGAGTGGCTGGAT | 891 |
| QY | 1189 | GGGACATCTTCAGACTACAAAAATTTGGAAGCTGGACAGCCCGATTAACCTGGGGTCATGGC | 1248 |
| Db | 892 | GGGACATCTTCAGACTACAAAAATTTGGAAGCTGGACAGCCCGATTAACCTGGGGTCATGGC | 951 |
| QY | 1249 | CATGGCCAGAGAGAGACTGTGCTGGGTGATTTATGCTGGGCAGTGGAAACGATTTTCAA | 1308 |
| Db | 952 | CATGGCCAGAGAGAGACTGTGCTGGGTGATTTATGCTGGGCAGTGGAAACGATTTTCAA | 1011 |
| QY | 1309 | TGTGAAGAGTCATAACTCTTATTCGGAAAAAGACAGGAGAACAGCT | 1355 |
| Db | 1012 | TGTGAAGAGTCATAACTCTTATTCGGAAAAAGACAGGAGAACAGCT | 1058 |

| ID | AAH43054 | standard; DNA; 2256 BP. |
|----|--|--------------------------------|
| AC | AAH43054; | |
| DT | 15-OCT-2001 | (first entry) |
| DE | Nucleotide sequence of a human scavenger receptor. | |
| KW | Human; scavenger receptor; SRCE-PI; macrophage; basal immunity; arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty | |
| KW | low density lipoprotein; ss. | |
| OS | Homo sapiens. | |
| FT | Key | Location/Qualifiers |
| FT | CDS | 74..1936 |
| FT | | /tag= a |
| FT | | /product= "scavenger receptor" |
| FN | WO200159107-A1. | |
| PD | 16-AUG-2001. | |
| PF | 08-FEB-2001; | 2001WO-JP00874. |
| PR | 14-FEB-2000; | 2000JP-0035155. |
| PR | 10-OCT-2000; | 2000JP-0309068. |
| PA | (FUSO) | FUSO PHARM IND LTD. |
| PI | Wakamiya N; | |

DR MPI: 2001-497076/54.
 DR P-PSDB: AAG63350.
 XX
 PT New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
 PT useful for treatment and diagnosis of diseases associated with oxidized
 PT low-density lipoprotein accumulation -
 XX
 PS Claim 4; Page 105-109; 118pp; Japanese.
 XX
 CC The present sequence encodes a human scavenger receptor, designated
 CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
 CC are useful in clarifying the functions of macrophages and basal
 CC immunity. They are also useful in the treatment, prevention, diagnosis
 CC and investigation of diseases such as arteriosclerosis, diabetic
 CC complications, bacterial infection and restenosis following angioplasty,
 CC which are associated with accumulation of oxidized low density
 CC lipoprotein and the binding of advanced glycation end-products into
 CC cells.
 XX
 XQ Sequence 2256 BP; 721 A; 524 C; 528 G; 483 T; 0 other;

| | | | | |
|-----------------------|-----------------|---------------------|-------------|--------------|
| query Match | 45.0%; | Score 617; | DB 22; | Length 2256; |
| Best Local Similarity | 72.9%; | Pred. No. 6.2e-151; | | |
| Matches 999; | Conservative 0; | Mismatches 0; | Indels 372; | Gaps 1, |

| | | | |
|----|------|--|------|
| QY | 1 | ATGAACGCCAGCTCAACATCTATTCACAGAGTCAGAGTGGAGAAATACATACCACTATCTCTAA | 60 |
| Db | 929 | ATGACACGCCAGCTCACTACTTATTCACAGAGTCAGAGTGGAGAAATACATACCACTATCTCTCAA | 988 |
| QY | 61 | GCCAAAGCAGCAGACTGTAAAGACCTGTGAGGACTTACACAAAGATGACAGAAATTAGAAC | 120 |
| Db | 989 | GCCAAACAGCAGAACTGTAAAGACCTGTGAGGACTTACACAAAGATGACAGAAATTAGAAC | 1048 |
| QY | 121 | GCCATCAAGTCAACCAACTGGAGGAAAGCCTTCACAGCTTTGTGACAGGATATTGTGAC | 180 |
| Db | 1049 | GCCATCAAGTCAACCAACTGGAGGAAAGCCTTCACAGCTTTGTGACAGGATATTGTGAC | 1108 |
| QY | 181 | ATCATTTAGCATATATCAAGTTACACAGGCCACCACTGTGGAGCCTGACACGACATCTAAAT | 240 |
| Db | 1109 | ATCATTTAGCATATATCAAGTTACACAGGCCACCACTGTGGAGCCTGACACGACATCTAAAT | 1168 |
| QY | 241 | GAACTCAGACCACTTGCACACATACCTTACCAAACACACAGATGATCTGACTCTTG | 300 |
| Db | 1169 | GAACTCAGACCACTTGCACACAGATACCTTACCAAACACACAGATGATCTGACTCTTG | 1228 |
| QY | 301 | AATATATCCCTGGCCAAACATCCGTTTGGATTCGTTCCTAGAGATCAACAAATTTG | 360 |
| Db | 1229 | AATATATCCCTGGCCAAACATCCGTTTGGATTCGTTCCTAGAGATCAACAAATTTG | 1288 |
| QY | 361 | ATGAGTCGAGTTTGAACACTGAAAGTAGGCCAACTTATCAGTGTATTATGGAAGAAATGAG | 420 |
| Db | 1289 | ATGAGTCGAGTTTGAACACTGAAAGTAGGCCAACTTATCAGTGTATTATGGAAGAAATGAG | 1348 |
| QY | 421 | CTATATACATCCCAAGCATGGTACGTCATCAACAAGATTTTACAATATCTCAAGAGTCCACCG | 480 |
| Db | 1349 | CTATATACATCCCAAGCATGGTACGTCATCAACAAGATTTTACAATATCTCAAGAGTCCACCG | 1408 |
| QY | 481 | GGCCCCAGGGGTCCCAAGAGGTGCACAAGAGATCCCAAGGAGCCCCCTGGGCCCAACTGGCAAC | 540 |
| Db | 1409 | GGCCCCAGGGGTCCCAAGAGGTGCACAAGAGATCCCAAGGAGCCCCCTGGGCCCAACTGGCAAC | 1468 |
| QY | 541 | AAGGACACAGAAAGAGAGAAAGGGGAGCCTGGACCACTGGCCCTTGGGGGTGACAGAGGC | 600 |
| Db | 1469 | AAGGACACAGAAAGAGAGAAAGGGGAGCCTGGACCACTGGCCCTTGGGGGTGACAGAGGC | 1516 |
| QY | 601 | CAAATTGGACACACTGGTCCCCCGGAGAGAGTGGGCGCAAGAGATCTAAAGAGCTCCAG | 660 |
| Db | 1517 | CAAATTGGACACACTGGTCCCCCGGAGAGAGTGGGCGCAAGAGATCTAAAGAGCTCCAG | 1516 |
| QY | 661 | GGCCCCAAAGAGCTCCCGTGGTTCCCTGTGGAAAGCCCGGCTTCAGAGGCCCAAGTGGGAGC | 720 |
| Db | 1517 | GGCCCCAAAGAGCTCCCGTGGTTCCCTGTGGAAAGCCCGGCTTCAGAGGCCCAAGTGGGAGC | 1516 |

QY 721 CCAAGCCCCCGGCCCCAGCAGCAAGAGGAGCTCCCGGCCCTCTCAGGGCCCTCTGCG 780
Db 1517 ----- 1516
QY 781 TTCAGGAGACTTCAGGCGACCGTTGGGAGCCTTGGGCTGCTGAGACTCGGGAGACTGCCA 840
Db 1517 ----- 1516
QY 841 GCGTTCGCTGGGGTACCAGGCATGCCAGGCCCCAGAGGGCCCCCGGCCCTCTCGGCCCA 900
Db 1517 ----- 1516
QY 901 TCAGAGAGCGGTGGTCCCTTGGCCCTGCGAGATGAGCCAAACCCCGGACCGGAGAGCAAT 960
Db 1517 ----- 1516
QY 961 GGGTCCCCGCTCAGTGAAGAAGTTCAGACAGCAATGCTACTATTTTTCAGTTGAGAAA 1020
Db 1517 GGGTCCCCGCTCAGTGAAGAAGTTCAGACAGCAATGCTACTATTTTTCAGTTGAGAAA 1576
QY 1021 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTGTGTTGATA 1080
Db 1577 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTGTGTTGATA 1636
QY 1081 AACACTGAGAGAGACAGCAATGATTAATAAAGACAGATGGTAGGAGAGAGAGCACTGG 1140
Db 1637 AACACTGAGAGAGAGCAATGATTAATAAAGACAGATGGTAGGAGAGAGAGCACTGG 1696
QY 1141 ATCGGCTCTCAGACTCAGAGCGTGAATAATGAAATGAGTGGCTGATGGGACATCTCCA 1200
Db 1697 ATCGGCTCTCAGACTCAGAGCGTGAATAATGAAATGAGTGGCTGATGGGACATCTCCA 1756
QY 1201 GACTACAAAAATTTGGAAGCTGACAGCCGAGTAACTGGGGTCATGGCCATGGGCCAGGA 1260
Db 1757 GACTACAAAAATTTGGAAGCTGACAGCCGAGTAACTGGGGTCATGGGGTCATGGGCCAGGA 1816
QY 1261 GAGAGCTGCTGGGTGATTTATGCTGGGCGAGTGGAGCAATTTCCATATGTAAGAGCTC 1320
Db 1817 GAGAGCTGCTGGGTGATTTATGCTGGGCGAGTGGAGCAATTTCCATATGTAAGAGCTC 1876
QY 1321 AATTAATCTTTCGCAAAAAAGACAGAGAGAGAGTACTGATCTGCAATTA 1371
Db 1877 AATTAATCTTTCGCAAAAAAGACAGAGAGAGAGTACTGATCTGCAATTA 1927

RESULT 14
AAAX27858
ID AAAX27858 standard; DNA; 3685 BP.
XX
XX AAAX27858:
AC
XX
XX 02-JUN-1999 (first entry)
DT
XX
XX Human CSR3 protein coding sequence.
DE
XX
XX Cellular stress response protein: CSR1; CSR2; CSR3; human; macrophage;
KW scavenger receptor protein; intracellular stress; arteriosclerosis;
KW diabetic circulatory obstruction; microbial infection; ss.
XX
XX Homo sapiens.
OS
XX
XX W09090159-A1.
PN
XX
XX 25-FEB-1999.
PD
XX
XX 12-AUG-1998; 98MO-JP03602.
PF
XX
XX 30-JUL-1998; 98JP-0230121.
PR
XX 13-AUG-1997; 97JP-0233396.
XX
XX (NIBS) JAPAN TOBACCO INC.
PA
XX
XX Nakamura Y, Tokyo T.
PI

XX
DR WPI; 1999-181032/15.
DR P-SDS; AAY00994.
XX
XX Scavenger receptor proteins - for treatment and diagnosis of
PT disorders involving cell stress
XX
XX
PS Claim 6; Page 142-150; 175pp; Japanese.
XX
XX This sequence encodes the human cellular stress response 3 (CSR3) protein
CC of the invention. The CSR proteins are macrophage scavenger receptor
CC proteins. The CSR proteins can be used in the treatment, gene therapy
CC and diagnosis of diseases in which intracellular stress is important,
CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial
CC infection. Expression of the proteins is induced in vivo in response to
CC intracellular stress, and inhibits cell death as a result of such stress.
XX
XX
SQ Sequence 3685 BP; 901 A; 1142 C; 972 G; 669 T; 1 other:

Query Match 17.2%; Score 235.6; DB 20; Length 3685;
Best Local Similarity 53.9%; Pred. No. 3,7e-51;
Matches 484; Conservative 0; Mismatches 414; Indels 0; Gaps 0;

QY 9 CCAGCTCACTCATTCACAGTGCAGATGAGAGAACATACCATATCTCTAAGCCCAACGA 68
Db 1056 CCAGCTCATGCGCTTGCAGCTGCAGTGAATACATCTCTCTCTGATGACCAACGA 1115
QY 69 GCAGAACCTTGAAGACCTCGAGACTTACACAAAGATGCAGAGATAGAACACCATCAA 128
Db 1116 AGAGAACATCATGATTTTACATACCATACCATACCATACCATACCATACCATACCAT 1175
QY 129 GTTCAACCAATGAGAGAACCTTCCAGCTTTTGAAGAGGATTTGTGACATCATTAG 188
Db 1176 GTTCAACCAATGAGAGAACCTTCCAGCTTTTGAAGAGGATTTGTGACATCATTAG 1235
QY 189 CAATATGATTTACAGAGCCCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 248
Db 1236 CAATATGATTTACAGAGCCCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1295
QY 249 GACCACTTGGCAGATATACCTTACCAAAACAGAGATGATCTGATCTCTTGAATAATAC 308
Db 1296 GCTCTCTGACAGCTGGGCTTCCACACCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1355
QY 309 CTTGGCCACATCCGTTTGGATTTCTGTTCTCAGAGATGCAACAAAGATTTGATGAGTC 368
Db 1356 TGTCTCTCATCATCTGAGGACCAACAGCCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1415
QY 369 GAGGTAGACAGTGAAGTACCAACTTATCAGATTAATGAGAAATGAAGTACAGTACGA 428
Db 1416 CCGGCTGGAGCTCAACGTCGGGAACCTTCTCATGATGCTGAGAGAGATGAGGAGTACGGA 1475
QY 429 CTCACAGCATGTCAGCTCATCAAGATTTTACATATCTCAAGAGTTCACCGGCCCCAG 488
Db 1476 CACACAGCATGAGAAATCTTGGCAATGTCACATCTCTGAGAGATGAGGAGTACGGA 1535
QY 489 GGGTCCAGAGGTGACAGAGATCCAGGAGACCCCTGCGCCCAACTGCAACAGGAGACA 548
Db 1536 AGGACCAAGAGATTAAGAGAGATATGGGCGCTGTTGGCGGAGGAGGAGGAGGAGGAGGAG 1595
QY 549 GAAAGAGAGAGAGGAGGAGCTGAGACCACTGCGCCCTGCGGAGTGAAGAGGCCAATTGG 608
Db 1596 GAAAGAGAGAGAGGAGGAGCTGAGACCACTGCGCCCTGCGGAGTGAAGAGGCCAATTGG 1655
QY 609 ACCAGCTGTCCTCCCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 668
Db 1656 AGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1715
QY 669 AGGCTCCGCTGTTCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 728
Db 1716 AGGCTCCGCTGTTCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1775
QY 729 CCGGAG 788
Db 788 CCGGAG

Db 1776 CATAGGCCCCCAGGAGGAGGCCCCGGGCTCTCCAGGGCCCTCAGGGCTTCAGGG 1835
 QY 789 ACTTCAGGGCACCCTTGGGAGCCTGGGGTGCCTGGACCTGGGGAGCTGCAGGGCTTGGC 848
 Db 1836 AAAACCCGGGAATTCAGGAGAGACAGGCTCACCAGGCCAGCGGGGGCCATGGGGCTTAA 1895
 QY 849 TGGGGTACCAGGCGATGCCAGGCCCCCGGCGCCCTCTGCGCCATCAGGA 906
 Db 1896 GGGTGAACAGGAGATCCAGGGTCCCGCTGTCTCCGGGGCCCTCCAGTCCACAGGA 1953

RESULT 15

AAx27856

ID AAx27856 standard; DNA; 3810 BP.

XX AAX27856;

XX 02-JUN-1999 (first entry)

XX Human CSRI protein coding sequence.

XX Cellular stress response protein; CSRI; CSRI2; CSRI3; human; macrophage;
 XX scavenger receptor protein; intracellular stress; arteriosclerosis;
 XX diabetic circulatory obstruction; microbial infection; ss.

XX Homo sapiens.

XX PM W09090159-A1.

XX 25-FEB-1999.

XX PD 12-AUG-1998; 98W0-JP03602.

XX PF 30-JUL-1998; 98JP-0230121.

XX PR 13-AUG-1997; 97JP-0233396.

XX PA (NISR) JAPAN TOBACCO INC.

XX PI Nakamura Y, Tokino T;

XX DR WPI; 1999-181032/15.

XX P-PSDB; AAY00992.

XX Scavenger receptor proteins - for treatment and diagnosis of
 XX disorders involving cell stress

XX PS Claim 4; Page 119-127; 175pp; Japanese.

XX This sequence encodes the human cellular stress response 1 (CSRI) protein
 XX of the invention. The CSRI proteins are macrophage scavenger receptor
 XX proteins. The CSRI proteins can be used in the treatment, gene therapy
 XX and diagnosis of diseases in which intracellular stress is important,
 XX such as arteriosclerosis, diabetic circulatory obstruction, and microbial
 XX infection. Expression of the proteins is induced in vivo in response to
 XX intracellular stress, and inhibits cell death as a result of such stress.

XX Sequence 3810 BP; 892 A; 1213 C; 1049 G; 655 T; 1 other;

Query Match 17.2%; Score 235.6; DB 20; Length 3810;

Best Local Similarity 53.9%; Pred. No. 3.8e-51;

Matches 484; Conservative 0; Mismatches 414; Indels 0; Gaps 0;

QY 9 CGAGCTCACTCATCTCACAGGTGATGAGAAATCACCACCTATCTCAAGCCACAGA 68
 Db 1181 CGAGGTATGGCTTGACAGCTGAGTGAATACATCTGCTCTCTGATGATCCAGCA 1240
 QY 69 GCAGAACTTGAAGACCTGACGAGATTACACAAGATGACAGAAATAGAACCCATCAA 128
 Db 1241 AGGAAGACATGATGATCTTCACTACCATACCACTACGCCAGAACCCATCTGAGAG 1300
 QY 129 GTTCAACCACTGAGGAAGAGCTTCCAGCTCTTTGAGACGATATTGTGAACATCATTA 188
 Db 1301 GTTTGAGTCTGTGAAGAGCATGGCTTCTACAGAGATTGAATTTGGACCATCTTAC 1360

QY 189 CAATATCACTTACACAGCCACACCTGCGAGAGCTGACACCAATCTAATGAATCAG 248
 Db 1361 CAATATCACTTACACAGCCACACCTGCGAGAGCTGACACCAATCTAATGAATCAG 1420
 QY 249 GACCACTTGCACAGATATCCCTTACCAACACACAGATGATCTGACCTCTTGAATATAC 308
 Db 1421 GCTCTCTCTGACAGCTGAGGCTTCCACACCATGCGGAGAGCTCTACTACTACCAAGTTC 1480
 QY 309 CCTGGCCAAATCTCCGTTTGGATTTCTTCTCTCAGAGATGCAACAAGATTTGATAGGTC 368
 Db 1481 TGTCTCTCATATGCTGGGAGCCACAGACCTGCTCCGGGAGCGCTTACGCTGTCACTGC 1540
 QY 369 GAGGTAGACACTGAATGAGCCACTATCAGATATTAAGAGAAATAGAGTAGTAGA 428
 Db 1541 CCGGCTGAGCTCAACAGCTCCGAACTCTCATGATGCTGAGAGATGAAGAGAGTGA 1600
 QY 429 CTCACAGCATGAGTGCAGCTCATCAAGAAATTTTCAATACTACAGAGTCCACCGGCCCCAG 488
 Db 1601 CACACAGCATGAGAGAAATCTCTGCAATGTCAACATCTACAGAGTCCCGGCCCTTCC 1660
 QY 489 GGGTCCAGAGGTGACAGAGATCCAGGAGCCCTGAGCCCAACTGCGCAACAGGAGACA 548
 Db 1661 AGGACCAAGAGGATTCAAAGAGATATGAGCGGTGAAGAGGCGCTGTGCGGAGAGGCC 1720
 QY 549 GAAAGSAGAGAGGGGAGGAGCTGAGCCACTGAGCCCTGAGGAGAGAGAGGCCAATTGG 608
 Db 1721 GAAAGSAGAGAGGGGAGGAGCTGAGCCACTGAGCCCTGAGGAGAGAGAGGCCAATTGG 1780
 QY 609 ACCAGCTGTCCCTCCCGGAGAGAGCTGAGGAGCAAGAGATCTAAAGGCTCCAGGAGCCCAA 668
 Db 1781 AGAGGCGGGGCTGTGAGAGAAAGAGGAGGCGCTGTGAGGCTTCCAGAGGCTTCAA 1840
 QY 669 AGGCTCCCTGTGTTCCCTGAGAGAGCGGAGCCCTCAAGGAGCCCAAGTGGGAGAGCCAGGCC 728
 Db 1841 AGGCTCAAGAGGAGCTTGTGAACCTGAGAGGAGGAGAGAGAGCCCAAGAGGAGGGA 1900
 QY 729 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 788
 Db 1901 CATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1960
 QY 789 ACTTCAGGACACCTTGGGAGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTTGGC 848
 Db 1961 AAAACCGGGAATTTGACAGGAGAAAGACAGGATCACAGGCCAGCGGGGGCCATGGGCTTAA 2020
 QY 849 TGGGATACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 906
 Db 2021 GGGTGAACAGGAGATCCAGGGTCCCGCTGTCTCCCGGGGCGCTCCAGGTCCACAGGA 2078

Search completed: March 21, 2003, 02:19:04

Job time : 298.589 secs

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 02:00:05 ; Search time 48.326 Seconds
(without alignments)
8700.356 Million cell updates/sec

Title: US-09-763-712A-1_COPY_325_1695

Perfect score: 1371

Sequence: 1 atgaacagccagctcaactc.....cagctactgtcatctgcattta 1371

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PTOTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------------------------|
| 1 | 121.6 | 8.9 | 810 | 1 | US-08-642-255-60 Sequence 60, Appl |
| 2 | 118 | 8.6 | 756 | 1 | US-08-642-255-50 Sequence 50, Appl |
| 3 | 113.6 | 8.3 | 3181 | 1 | US-08-655-086-1 Sequence 1, Appl |
| 4 | 112.8 | 8.2 | 506 | 1 | US-08-253-155A-61 Sequence 61, Appl |
| 5 | 106 | 7.7 | 432 | 1 | US-08-642-255-48 Sequence 48, Appl |
| 6 | 106 | 7.7 | 1608 | 4 | US-09-029-348-19 Sequence 19, Appl |
| 7 | 105.4 | 7.7 | 503 | 1 | US-08-253-155A-20 Sequence 20, Appl |
| 8 | 100 | 7.3 | 5102 | 1 | US-08-494-168-1 Sequence 1, Appl |
| 9 | 99.6 | 7.3 | 1416 | 2 | US-07-621-091G-1 Sequence 1, Appl |
| 10 | 99.6 | 7.3 | 1416 | 2 | US-08-399-889-1 Sequence 1, Appl |
| 11 | 99.6 | 7.3 | 1416 | 2 | US-09-167-364-1 Sequence 1, Appl |
| 12 | 99.6 | 7.3 | 1416 | 4 | US-09-439-897-1 Sequence 1, Appl |
| 13 | 98.4 | 7.3 | 1868 | 1 | US-08-392-367B-1 Sequence 1, Appl |
| 14 | 98.4 | 7.3 | 1868 | 3 | US-08-893-467A-1 Sequence 1, Appl |
| 15 | 99.4 | 7.3 | 2543 | 3 | US-08-555-669-11 Sequence 11, Appl |
| 16 | 99.4 | 7.3 | 2543 | 3 | US-09-073-663-11 Sequence 11, Appl |
| 17 | 97.6 | 7.1 | 4359 | 4 | US-09-484-970B-4 Sequence 4, Appl |
| 18 | 96.2 | 7.0 | 1572 | 4 | US-09-297-269-39 Sequence 39, Appl |
| 19 | 95.2 | 6.9 | 1560 | 2 | US-08-794-795-5 Sequence 5, Appl |
| 20 | 95.2 | 6.9 | 1560 | 4 | US-09-249-200-5 Sequence 5, Appl |
| 21 | 95.2 | 6.9 | 1703 | 2 | US-08-794-795-1 Sequence 1, Appl |
| 22 | 95.2 | 6.9 | 1703 | 4 | US-09-249-200-1 Sequence 1, Appl |
| 23 | 89 | 6.5 | 1881 | 4 | US-09-029-348-20 Sequence 20, Appl |
| 24 | 77 | 5.6 | 1560 | 4 | US-09-453-702B-264 Sequence 264, Appl |
| 25 | 77 | 5.6 | 61663 | 4 | US-09-453-702B-62 Sequence 62, Appl |
| 26 | 76.8 | 5.6 | 5552 | 3 | US-08-155-888-1 Sequence 1, Appl |
| 27 | 74.6 | 5.4 | 38584 | 4 | US-09-453-702B-50 Sequence 50, Appl |

| | | | | | |
|----|------|-----|-------|---|---------------------------------------|
| 28 | 73 | 5.3 | 503 | 4 | US-09-297-269-40 Sequence 40, Appl |
| 29 | 72.4 | 5.3 | 4031 | 1 | US-08-159-784-1 Sequence 1, Appl |
| 30 | 72.2 | 5.3 | 9827 | 4 | US-09-453-702B-66 Sequence 66, Appl |
| 31 | 71.4 | 5.2 | 3927 | 4 | US-09-404-879A-372 Sequence 372, Appl |
| 32 | 70 | 5.1 | 513 | 4 | US-09-134-001C-647 Sequence 647, Appl |
| 33 | 70 | 5.1 | 585 | 4 | US-09-134-001C-726 Sequence 726, Appl |
| 34 | 70 | 5.1 | 810 | 4 | US-09-134-001C-624 Sequence 624, Appl |
| 35 | 70 | 5.1 | 3552 | 4 | US-09-134-001C-693 Sequence 693, Appl |
| 36 | 69 | 5.0 | 1347 | 4 | US-09-140-804-1 Sequence 1, Appl |
| 37 | 68.6 | 5.0 | 45175 | 4 | US-09-453-702B-116 Sequence 116, Appl |
| 38 | 67.8 | 4.9 | 3404 | 4 | US-09-453-702B-94 Sequence 94, Appl |
| 39 | 67.8 | 4.9 | 48908 | 4 | US-09-453-702B-137 Sequence 137, Appl |
| 40 | 67.2 | 4.9 | 729 | 4 | US-09-140-804-10 Sequence 10, Appl |
| 41 | 67.2 | 4.9 | 226 | 1 | US-08-175-155-63 Sequence 63, Appl |
| 42 | 67 | 4.9 | 226 | 1 | US-08-477-509B-98 Sequence 98, Appl |
| 43 | 67 | 4.9 | 226 | 1 | US-08-642-255-96 Sequence 96, Appl |
| 44 | 67 | 4.9 | 226 | 1 | US-08-397-633A-16 Sequence 16, Appl |
| 45 | 67 | 4.9 | 226 | 2 | US-08-707-237A-70 Sequence 70, Appl |

ALIGNMENTS

RESULT 1
US-08-642-255-60
Sequence 60, Application US/06642255
Patent No. 5773249
GENERAL INFORMATION:
APPLICANT: CAPPELLO, Joseph
APPLICANT: FERRARI, Franco A.
TITLE OF INVENTION: High Molecular Weight Collagen-Like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A55556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 810 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLBUE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic"
US-08-642-255-60
Query Match 8.9%; Score 121.6; DB 1; Length 810;
Best Local Similarity 53.9%; Pred No. 4.1e-23;
Matches 250; Conservative 0; Mismatches 214; Indels 0; Gaps 0;
QY 465 ACTACAAGTCCACCGCGGCCGAGGGGTCCAGAGGTGACGAGATCCAGGACCC 524
|| || ||| |||| ||||| ||||| || ||||| ||

Db 210 ACCACCGGGTGCACCTGGCCACGCGGTCCGCTGCATCTAGAGGTGACCCAGACCGCC 269

QY 525 TGGCCCACTGGCAACAGGAGAGAAAGGAGAGAAAGGGAGCCTGGACCTGAGCC 584

Db 270 TGGTGGTCCGGGACCTGACAGGACCGGCTAGCCGTGGGAGATCCGGAGCCACCGGGTGC 329

QY 585 TGGGGGTGAGAGAGGCCCAATTGGACAGCTGTCCCGGAGAGAGCTGGCGCAAGG 644

Db 330 ACTTGCGCCACGCGGGTCCGCTGTGATCTAGAGGTGACCGGAGCAGCGCTGTGCTCCGG 389

QY 645 ATCTAAGGCTCCAGGGGCCCCAAAGGCTCCGCTGGTTCCTGGAGAGCCCGGCTCA 704

Db 390 ACTGCAAGGCCCCACCGGATGAGCTGTGCTGCGGAGACACCGGATGACCTGGCCAGC 449

QY 705 GGGCCCACTGGGAGCCAGGCCCCCGGGCCACAGGCAAGAGGAGTCCCGGCCC 764

Db 450 GGGTCCGCTGGATCTAGAGGTGACCGGAGCCGCTGGTGTCCGGAGCTGAGAGCC 509

QY 765 TCAAGGCCCCCTGCTGCTTCCAGGACTTCAAGGACCGCTTGGGAGCTGGGCTGCTGG 824

Db 510 ACCGGGTAGCCGCTGGGATCCGGGACCCGCGGTGACACCTGGCCACAGCGGCTCGCTGG 569

QY 825 ACCGCGGAGCTGCAGGCTTGGCTGGGATACAGGATGCAAGGCCCCCAAGGGCCCC 884

Db 570 ATCTAGAGGTGACCCAGGACCGCTGTGCTCCGGAGCTGCAAGGCCACCGGGTACGCG 629

QY 885 CGGCCCCCTGCGCCATCAGAGCGGTGTGCCCCCTGGCCCTGC 928

Db 630 TGGGATCCGGGACACAGCGGGTGCACCTGGCCACAGCGGGTCCCG 673

RESULT 2

US-08-642-255-50

Sequence 50, Application US/08642255

Patent No. 5773249

GENERAL INFORMATION:

APPLICANT: CARPELLO, Joseph

APPLICANT: FERRARI, Franco A.

TITLE OF INVENTION: High Molecular Weight Collagen-Like

TITLE OF INVENTION: Protein Polymers

NUMBER OF SEQUENCES: 135

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/642,255

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: ROWLAND, Berttram I.

REGISTRATION NUMBER: 20,015

REFERENCE/DOCKET NUMBER: A55556-3/BIR

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 494-8700

TELEFAX: (415) 494-8771

TELEX: 910 277299 FHT UR

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:

LENGTH: 756 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "synthetic"

US-08-642-255-50

Query Match

Best Local Similarity 8.6%; Score 118; DB 1; Length 756;

Matches 238; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 472 GGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGATCCAGAGGACCCCTGGCCCA 531

Db 289 GGTCCACCGGGTCTCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 348

QY 532 ACTGGCAACAAG 591

Db 349 CCGGGGTCTCCGGAGACTGACAGGCCCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 408

QY 592 GAGAGAGGCCCCCAATTGGAGGAGTGTCTCCCGGAGAGGAGGAGGAGGAGGAGGAGGAG 651

Db 409 GCTCCGGGAGCTCAGAGGCCCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 468

QY 652 GGTCCACCGGGCCCCAAAGGCTCCGCTGTGCTTCCCTGGAAGAGGAGGAGGAGGAGG 711

Db 469 GGAAGCTGCAAGGCCCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 528

QY 712 AGTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 771

Db 529 GCAAGGCCCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 588

QY 772 CTTCTGAGGCTTCCAGGAGGAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 831

Db 589 CCGCCAGAGTGGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 648

QY 832 GGAAGTCCAGGCTTGGCTGGGAGTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 891

Db 649 GGTGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 708

QY 892 CTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 909

Db 709 CATGGCCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 726

RESULT 3

US-08-655-086-1

Sequence 1, Application US/08655086

Patent No. 5821089

GENERAL INFORMATION:

APPLICANT: GROSCHN, ELLIOTT A.

APPLICANT: BUECHTER, DOUGLAS

APPLICANT: ZHANG, GUANGHUI

APPLICANT: CONNOLLY, KEVIN

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: DILWORTH & BARRESE

STREET: 333 EARLE O'VINGTON BOULEVARD

CITY: UNIONDALE

STATE: NY

COUNTRY: US

ZIP: 11553

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/655,086

FILING DATE: 03-JUN-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: STEEN, JEFFREY S.

REGISTRATION NUMBER: 32,063

REFERENCE/DOCKET NUMBER: 203-1632

TELECOMMUNICATION INFORMATION:

TELEPHONE: 516-228-8484

TELEFAX: 516-228-8516

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 3181 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cdna
US-08-655-086-1

Query Match 8.3%; Score 113.6; DB 1; Length 3181;
Best Local Similarity 52.5%; Pred. No. 1e-20;
Matches 248; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 442 CAGCTATCAAGATTTTCAATACATACAGAGTCCACCGGGCCCGAGGGTCCAAAGAGT 501
DB 104 CTGGTCCCAAGAGCTTCCCAAGGCTTCCAGAGTCCCGCTGGAGAGCTGCGAGCTGGA 163
QY 502 GACAGAGATCCAGAGGAGCCCGCTGCGCCCACTGGCAACAGGAGACGAAAGAGAGAG 561
DB 164 GCTTACAGGTCCTGGGTCCCGAGAGTCCCGCAGGTCCCGTGAAGAAAGATGATGAT 223
QY 562 GGGGAGCCCTGGACACCTGGCCCTGGGGTGAAGAGGCCCAATTGGACAGCTGTCCTG 621
DB 224 GGGGAGGCTGGAAACCTGGTCTGCTGGTGAAGCGTGGGCTCTCTGGGCTCAGGGTGT 283
QY 622 CCCGAGAGCTGGGGGCAAGAGATCTAAAGCTCCAGAGGCCCAAGAGCTCCCTGGT 681
DB 284 CGAGATTCGCCGGAACAGACTGGCCCTCCCTGGAATGAAGGACAGAGGTTTCACTGAT 343
QY 682 TCCCTGGAGAGCCCGGCTCAGGGCCCACTGGGGAGCCAGGCCCGCGGGCCACCA 741
DB 344 TTGGATGGTCCCAAGGAGATGTGTGCTGCTGCTGAAGGATGAGCTGGCAGCCCT 403
QY 742 GGCAAGAGGAGACTCCCGGCTCAGGGCCCTCTGGGCTTCCAGGAGATTACAGGCAC 801
DB 404 GTGAAATGAGCTCTGTGATGATGGCCCGCTGGCTGCTGTAAGAGAGGTGCG 463
QY 802 GTTGGGAGAGCTGGGCTGGCTGGAGCTCGGGAGACTCCAGGCTTGGGGGTACAGAGC 861
DB 464 CCGTGGAGCCCTGGGCTGCTGTGCTGTGGAATGATGATGCTGTGCTGTGCGGG 523
QY 862 ATCCAGAGCCCGAGGCCCCCGGCTCTGTGGCCATCAGAGCGGTGG 913
DB 524 CCGCTGGTCCACCGGCCCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 575

RESULT 4
US-08-235-155A-61/c
; Sequence 61, Application US/08253155A
; Patent No. 5691147

GENERAL INFORMATION:
APPLICANT: Gyuris, Jeno
APPLICANT: Draetta, Giulio
TITLE OF INVENTION: CDK4 Binding Proteins
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/253,155A
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709

REFERENCE/DOCKET NUMBER: MIT-028
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 506 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-08-253-155A-61

Query Match 8.2%; Score 112.8; DB 1; Length 506;
Best Local Similarity 53.8%; Pred. No. 7.5e-21;
Matches 254; Conservative 0; Mismatches 217; Indels 1; Gaps 1;

QY 182 TCATTAGCAATATCAGTTTACAGAGCCACCACTGGGAGCGTG-ACCAGCAATCAAT 240
DB 506 TCTTACCAACATCAATTCACCGAGCCACAGTGCACAAACATGCTCCAAAGTACCTGGAT 447
QY 241 GAAGTCAGACCACTTGCACAGATACCTTACCAACACACAGATGATGTGACTGCTTG 300
DB 446 GACGTTCGGCTCTCCGACAGCTGGGCTTCCACACCCATGCGAGAGGCTTACTACTG 387
QY 301 AATTAATCCCTGGCCCAACATCGGTTTGATTCGTTCTCAGAGATGCAACAAGATTG 360
DB 386 AACAGTGTCTCTCCATCATGCTGGGACACACAGACTCTCGGGAGGCTTCAAGCTG 327
QY 361 ATGAGTCTGAGGTTAGACACTGAGTACCACTTATCAGTATTAAGAAATGAG 420
DB 326 CACAGTCCCGGCTGAGCTCAACGTCGGAACTTCATGATGCTGAGGAGATGAAG 267
QY 421 CTAGTAGCTCCAAGCTGTGACGCTCATCAAGATTTTACAACTACTACAGGTCACCG 480
DB 266 GCAGTGAGACACAGCATGAGAAATCTTCGCAATGTCACATCTACAGAGTGCCCC 207
QY 481 GGGCCCAAGGGGTCAAGAGGTGACAGAGATCCAGAGGCCCGCTGGCCCACTGGCAAC 540
DB 206 GGCCCTCCAGAGACCAAGAGATTTCAAGAGATATGGGGGTGAAGAGGCTGTGGGGC 147
QY 541 AAGGACAGAAAGAGAGAGGAGGAGGAGCTGACCACTGGCCCTGCGGGTGAAGAGGC 600
DB 146 AGAGGCCGGAAGAGAGACCCCGGACGCTTGGGCCCCCTGGAGGCCAGGGCTCTCAAG 87
QY 601 CCAATTGACACAGCTGTCCTCCCGGAGAGCGTGGCGCAAGATCTAAG 652
DB 86 CAACCTGAGAGAGGCGGCGCTGTGGGAGAAAGGGCCCTGTGGCCCTCGAG 35

RESULT 5
US-08-642-255-48
; Sequence 48, Application US/08642255
; Patent No. 5773249

GENERAL INFORMATION:
APPLICANT: CAPPELLO, Joseph
APPLICANT: FERRARI, Franco A.
TITLE OF INVENTION: High Molecular Weight Collagen-Like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLUHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/642,255
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Berttram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A55556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 27299 FHT UR
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic"

Query Match 7.7%; Score 106; DB 1; Length 432;

Best Local Similarity 54.0%; Pred. No. 4.5e-19;
Matches 217; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 508 GATCCAGGAGGACCCCTGAGCCCACTGCGACAGGAGACGAAAGAGAGAGAGGAG 567
DB 1 GGTCTACACGCGCCAGAGAGTCCGAAAGGCGCGCATGCGCCAGAGAGCCGAAAGGTGCG 60
QY 568 CTTGGAGCCACTGCGCCCTGCGGGGTGAGAGAGGCCCAATTGAGCAGCTGTGTCCTCCCGGA 627
DB 61 CTTGGAGCCACTGCGCCCTGCGGGGTGAGAGAGGCCCAATTGAGCAGCTGTGTCCTCCCGGA 120
QY 628 GAGCCTGGCGGCGCAAGAGATTAAGGCTCCAGAGGCCCAAGAGCTCCCTGTTCCCT 687
DB 121 CCGGCTGTGCTACCGGGTGTCTCGGGAGCTGCAAGGCCCGCCAGTGGCTGTGACCGGCT 180
QY 688 GGGAGCCCGGCGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 747
DB 181 GGTCCACCGGCTCTCGGGAGCTGCGAGGCCCGCCAGAGTCCGCTGAGAGCGGTGTCGA 240
QY 748 GAGGAGACTCCCGGCGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 807
DB 241 CCGGCTGTGCTACCGGGTGTCTCGGGAGCTGCAAGGCCCGCCAGTGGCTGTGACCGGCT 300
QY 808 GAGCCTGGCGGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 867
DB 301 GGTCCACCGGCTCTCGGGAGCTGCGAGGCCCGCCAGAGTCCGCTGAGAGCGGTGTCGA 360
QY 868 GGGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 909
DB 361 GGACTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 402

RESULT 6
US-09-029-348-19

; Sequence 19, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: 087857/PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1608
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS

US-09-029-348-19

Query Match 7.7%; Score 106; DB 4; Length 1608;

Best Local Similarity 52.8%; Pred. No. 8e-19;
Matches 229; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 472 GGTCCACCGGCGCCCAAGGAGTGCAGAGATCCAGGAGCCCTGAGCCCA 531
DB 361 GGTGAACCTGCTCAAACTGCTCGAGGTCACCTGTCATGAGCCCGTGGGTCT 420
QY 532 ACTGCAACAAGGAG 591
DB 421 GGTGAACCTGCTCAAACTGCTCGAGGTCACCTGTCATGAGCCCGTGGGTCT 480
QY 592 GAGAGAGCCCAATTGAGCAGCTGTGTCCTCCCGAGAGAGAGAGAGAGAGAGATTA 651
DB 481 GCTGTGCGCCCAAGAGTCTAGTGGCCCAAGAGATTCGTGCGATTAAGAGAGAGCC 540
QY 652 GGTCCAGAGGCGCCCAAGAGCTCCGTTCCCTGGAGAGCCCGCCCTCAGGAGCC 711
DB 541 GGTGAACCTGCTCAAACTGCTCGAGGTCACCTGTCATGAGAGATTAAGAGCTG 600
QY 712 AGTGGAGAGCCAGGCGCCCGGCGCCAGAGCAAGAGAGAGAGAGAGAGAGATTC 771
DB 601 CTTGATGCTGTGACCATGATGATCAAGGTCCTGCTGCTGCTGCTGCTGCTGCT 660
QY 772 CTTCTGCTGCTTCCAGGAGTTCAGAGGACCTGTTGGGAGAGCTGGGCTGAGCTCG 831
DB 661 CTTAGAGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 832 GACTGCGAGGCTTGGCTGAGGATTCAGAGGATTCAGAGGATTCAGAGGATTCAGAGG 891
DB 721 GGTAGAGGTTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 892 CTTGCGCATCAGG 905
DB 781 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 794

RESULT 7
US-08-253-155A-20

; Sequence 20, Application US/08253155A
; Patent No. 5691147
; GENERAL INFORMATION:
; APPLICANT: Gyuris, Jeno
; APPLICANT: Diarella, Giulio
; TITLE OF INVENTION: CDK4 Binding Proteins
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/253,155A
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MIT-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:

LENGTH: 503 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 US-08-253-155A-20

Query Match 7.7%; Score 105.4; DB 1; Length 503;
 Best Local Similarity 52.9%; Pred. No. 7e-19;
 Matches 226; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 63 CACGAGCAGAACTGTAAGACCTGAGAGCTTACCAAAAGATGCAAGATAGAACAC 122
 Db 15 CCACGAGAGAACATGATGATCTTACATACCTACGCGCCAGAACCGCACTGT 74
 QY 123 CATCAAGTTCAACCAACGAGAGAGAGAGCTTCCAGCTCTTTGAGAGGATTTGTGAACAT 182
 Db 75 GGAGAGGTTTGAGTCTGTGTGAGAGCGATGGCTTCTCAGAGATTGAATTGGCCACT 134
 QY 183 CATTAAGCATATCAGTTACACAGCCACCTGCGAGCGTACCGAGCATCTAATATGA 242
 Db 135 CTTACACCAATCATATGCGACGACGACGACGACGACATGCTCATCTACTGATGA 194
 QY 243 AGTCAGGACCACTTGCACAGATATCCCTTACCAAAACACAGATGATCTGACCTTGA 302
 Db 195 CGTGGGGCTCTCTGACAGCTGGGCTTCCACACCCATGCGGAGGCTCTACTACTGA 254
 QY 303 TAATACCTGCGCCACATCCGTTTGGATTTCTTCTCAGAGATGCAACAAAGATTGAT 362
 Db 255 CAAGTCTGTCTCATATGCTGTGGGACCAACACCTGCTCCGGGAGCGGTTACAGCTGCT 314
 QY 363 GAGTCGAGAGTTAGACACTGAGATGAGCCAACTTATCAGTATTTAGGAATGAAAGCT 422
 Db 315 CAGTGGCCGGCTGGAGCTTCAAGCTCCGGAACCTTCTCATGATCGTGAAGATGAAGG 374
 QY 423 AGTAGATCCCAAGCATGTGCTCATCAAGATTTTACATATACAAAGTCCACCGG 482
 Db 375 AGGGGACACAGAAATGGGAGATCCTTGGAAATGAATCCTACGAGGTGCCCGCGC 434
 QY 483 CCCCAGG 489
 Db 435 CTCGAGG 441

RESULT 8

US-08-494-168-1
 Sequence 1, Application US/08494168

Patent No. 5731192

GENERAL INFORMATION:

APPLICANT: Readers, Stephen T.

APPLICANT: Zhou, Jing

TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method

TITLE OF INVENTION: of Detecting Collagen Deficiency

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/494,168

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/112,465

FILING DATE: 27-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: SAXE, Bernhard D.
 REGISTRATION NUMBER: 28,665
 REFERENCE/DOCKET NUMBER: 40397/104/BAHR
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5102 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: join(2..82, 86..97, 101..4399, 4403..4420, 4424

US-08-494-168-1

Query Match 7.3%; Score 100; DB 1; Length 5102;
 Best Local Similarity 52.4%; Pred. No. 5.3e-17;
 Matches 220; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 468 ACAAGTCCACCGGCGCCCGAGGGTCCAAAGAGTGAACAGGATCCACAGACCCCTGG 527
 Db 2380 AATAGTGTCTCCGGGGGAGAACAGGCTTACAGAGATTAAACAGGGCAAAAGATTCTTG 2439
 QY 528 CCCACTGGCAACAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 587
 Db 2440 AGACTCTGCTCTCCAGACTCAAGGGTGTGACAGGGAGAGAGAGAGAGAGAGAGAG 2499
 QY 588 GGGTGAG 647
 Db 2500 AGGTGAGCGGGGAG 2559
 QY 648 TAAAGGCTCCAGAGGCGCCCAAGAGCTCCGTTGTTCCCTGGGAAAGCCCGGCTTGA 707
 Db 2560 TATGTGTCATATGATGATCAAGGGCAATCTGGGCTCCAGAGAGAGAGAGAGAGAG 2619
 QY 708 CCCCAGTGGGAG 767
 Db 2620 CATCTCAGAGACATCTCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2679
 QY 768 GGGCCCTCTGCTTCCAGAGGACTTCCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 827
 Db 2680 TGTAAAGAAAGAGGCTGCGAGGCTTAAAGGCTTCTTGAATTCAGGCTTATGAGACT 2739
 QY 828 TCGGGAGCTGCCAGGCTTGTGAGGCTTACAGGAGCATCCAGGAGAGAGAGAGAGAG 887
 Db 2740 GAAAGGAG 2799

RESULT 9

US-07-621-091G-1
 Sequence 1, Application US/07621091G

Patent No. 5424408

GENERAL INFORMATION:

APPLICANT: Readers, Stephen T., Morrison, Karen E., Hudson, Billy

APPLICANT: G.

TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen

TITLE OF INVENTION: Polynucleotides

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: Yale University, Office of Cooperative Research

STREET: 246 Church Street

CITY: New Haven

STATE: Connecticut

COUNTRY: U.S.A.

ZIP: 06510

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 800K storage

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh OS7.0

```
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/621,091G
FILING DATE: 11/30/90
CLASSIFICATION: 424
PRIOR APPLICATION DATA: NO. 5424408 applicable
ATTORNEY/AGENT INFORMATION:
NAME: Barth, Richard S.
REGISTRATION NUMBER: 28180
REFERENCE/DOCKET NUMBER: 900983/RB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 972-1400
TELEFAX: (212) 370-1622
TELEX: 236268
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1416 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: CDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Calf
STRAIN: Unknown
INDIVIDUAL ISOLATE: Unknown
DEVELOPMENTAL STAGE: Unknown
CELL TYPE: Whole kidney
IMMEDIATE SOURCE:
LIBRARY: Bovine lens cDNA
CLONE: KMC15
POSITION IN GENOME: NO. 5424408 known
CHROMOSOME/SEGMENT: NO. 5424408 known
PUBLICATION INFORMATION: NO. 5424408e
US-07-621-091G-1
```

```
Query Match 7.3%; Score 99.6; DB 1; Length 1416;
Best Local Similarity 51.6%; Pred. No. 3.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;
```

```

472 GGTCCACCGGGCCCGGAGGTCGACAGGATCCCGGAGCCCGCTGGCCCA 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 GGAACACCAAGTGTCCCGGTCAAGGAGGAGGATCTGATCTATGATTT 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
532 ACTGGCAACAAGGACAAAGAGAGAGAGGAGCGCTGACCACTGCGGGT 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
343 CAGAGCATGAAGGAGAGAGGATTTCAAGATTTCAGAGACCACTGGAC 402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
592 GAGAGAGCCCAATTGACCAAGCTGCTCCCGGAGAGCGTGGCCCAAGATCTAA 651
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 CAAGTGTGACCAAAAGGACCACTGAGTACGTGAGAGCGTGGCACAGTAAGATATC 462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
652 GGTCCACCGGGCCCGGAGGTCGACAGGATCCCGGAGCCCGCTGGCCCA 711
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
463 TCCCTTCAGAGAACCCAGGCGCACTGTGTTCAAGTGAAGACAGGAGTGAAGGAA 522
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
712 AGTGGGAGCCAGGCGCCCGGGCCACACAGGCAAAAGAGGACTCCCGGCTCAAGGC 771
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
523 CCGGGGCCCCAGAGACCAAGAGATCCAGAACCTGTGGGCCAAAGGTTAAACAGGG 582
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
772 CTTCTGGCTTCCAGGAGCTTCAAGGACCGCTTGGGAGCGCTGGGCTGCTGACCTGG 831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
583 GAGATGTGTCACACAGAACTCTGTGACCACTGAGCAAAAGGCAAAAGGTTGAAA 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
832 GGAAGTGCAGAGCTTGGCTGGGGTACACAGGATCCAGGCCCAAGGGCCCCCGGCGCT 891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
643 GGAAGGCAAGGACCACTGATCCGATGCTGCTGACAGCTTTGAAGGGGAAACCTGGAGAC 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
892 CTTGGCCCATCAGAGAGCGGTGG 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
703 ACTGGACCACTGCAGCAGGGG 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 10
US-08-399-889-1
Sequence 1, Application US/08399889B
Patent No. 5973120
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263A
CURRENT APPLICATION NUMBER: US/08/399,889B
CURRENT FILING DATE: 1995-03-07
EARLIER APPLICATION NUMBER: 07/621091
EARLIER FILING DATE: 1990-11-30
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1
LENGTH: 1416
TYPE: DNA
ORGANISM: Calf
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1416)
US-08-399-889-1
```

```
Query Match 7.3%; Score 99.6; DB 2; Length 1416;
Best Local Similarity 51.6%; Pred. No. 3.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;
```

```

472 GGTCCACCGGGCCCGGAGGTCGACAGGATCCCGGAGCCCGCTGGCCCA 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 GGAACACCAAGTGTCCCGGTCAAGGAGGAGGATCTGATCTATGATTT 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
532 ACTGGCAACAAGGACAAAGAGAGAGAGGAGCGCTGACCACTGCGGGT 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
343 CAGAGCATGAAGGAGAGAGGATTTCAAGATTTCAGAGACCACTGGAC 402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
592 GAGAGAGCCCAATTGACCAAGCTGCTCCCGGAGAGCGTGGCCCAAGATCTAA 651
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 CAAGTGTGACCAAAAGGACCACTGAGTACGTGAGAGCGTGGCACAGTAAGATATC 462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
652 GGTCCACCGGGCCCGGAGGTCGACAGGATCCCGGAGCCCGCTGGCCCA 711
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
463 TCCCTTCAGAGAACCCAGGCGCACTGTGTTCAAGTGAAGACAGGAGTGAAGGAA 522
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
712 AGTGGGAGCCAGGCGCCCGGGCCACACAGGCAAAAGAGGACTCCCGGCTCAAGGC 771
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
523 CCGGGGCCCCAGAGACCAAGAGATCCAGAACCTGTGGGCCAAAGGTTAAACAGGG 582
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
772 CTTCTGGCTTCCAGGAGCTTCAAGGACCGCTTGGGAGCGCTGGGCTGCTGACCTGG 831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
583 GAGATGTGTCACACAGAACTCTGTGACCACTGAGCAAAAGGCAAAAGGTTGAAA 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
832 GGAAGTGCAGAGCTTGGCTGGGGTACACAGGATCCAGGCCCAAGGGCCCCCGGCGCT 891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
643 GGAAGGCAAGGACCACTGATCCGATGCTGCTGACAGCTTTGAAGGGGAAACCTGGAGAC 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
892 CTTGGCCCATCAGAGAGCGGTGG 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
703 ACTGGACCACTGCAGCAGGGG 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 11
US-09-167-364-1
Sequence 1, Application US/09167364
Patent No. 6007980
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263B
```


;; CURRENT APPLICATION NUMBER: US/09/167,364
;; CURRENT FILING DATE: 1998-10-07
;; EARLIER APPLICATION NUMBER: 08/399889
;; EARLIER FILING DATE: 1995-03-07
;; NUMBER OF SEQ ID NOS: 25
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 1416
;; TYPE: DNA
;; ORGANISM: calf
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(1416)
US-09-167-364-1

Query Match 7.3%; Score 99.6; DB 3; Length 1416;
Best Local Similarity 51.6%; Pred. No. 3.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 472 GGTCCACCGGCCCCAGGGGTCACAGAGGATCCAGGGAGCCCCCTGGCCCA 531
DB 283 GGAGCCACAGGTGTCCCGGTACAGAGGGCCAGAGGTGATCTTGATTT 342
QY 532 ACTGGCAACAGGACAGAAAGAGAAAGGGGAGCTTGACACCTGGCCCTGGGGT 591
DB 343 CCAGGATGAAGGAGAAAGGTAATTCAGATTTCCAGGACACCTGGACCTCCAGGG 402
QY 592 GAGAGAGGCCCAATTTGACACAGTGTCCCGGAGAGAGCTGGCGGCAAGATCTAA 651
DB 403 CAAGTGGACCAAAAGACACCTGATGACGTGAGAGAGCTGGCAGTGAAGATC 462
QY 652 GGCTCCACAGGCCCCCAAGAGCTCCGTGGTCCCTGGGAGAGCCCGCCCTCAGGGCCC 711
DB 463 TCCCTTCAGAGAAAGCCAGGCCACCTGGTTACGTGAGAAACAGGATGACAGAGAA 522
QY 712 AGTGGGAGCCAGGCCCCCGGGCCACAGGCAAGAGAGTCCCGCCCTCAGGGC 771
DB 523 CCGGGGCCCCAGAGACACAGAGATCCAGGACCTGTGGGCAAAAGGTAACAGGG 582
QY 772 CCTCTGGCTTCAGGAGACTTCAGGGCACCCTGGGGAGCTGGGGTGCCTGACCTGG 831
DB 583 GAGGATGTTCACAGCACTCCTGGACCACTGGAGAAAGCAAAAGGTTGTAA 642
QY 832 GGACTCCAGAGCTTGGCTGGGGTACAGGATGCCAGGCCCCAGGGCCCCCGCT 891
DB 643 GGAGAGCAAGAGACACCTGATGATGCTGCTGCCAGGCTTGAAGGGAAACCTGAGAC 702
QY 892 CCTGGCCATCAGAGCGGTGG 913
DB 703 ACTGGACACCTGACAGCAGGG 724

RESULT 12
US-09-439-897-1

;; Sequence 1, Application US/09439897
;; Patent No. 6277558
;; GENERAL INFORMATION:
;; APPLICANT: Hudson, Billy G
;; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
;; FILE REFERENCE: 95-1263-C
;; CURRENT APPLICATION NUMBER: US/09/439,897
;; CURRENT FILING DATE: 1999-11-12
;; NUMBER OF SEQ ID NOS: 65
;; SOFTWARE: Patentln Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 1416
;; TYPE: DNA
;; ORGANISM: Bos taurus
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(1416)
US-09-439-897-1

Query Match 7.3%; Score 99.6; DB 4; Length 1416;
Best Local Similarity 51.6%; Pred. No. 3.9e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 472 GGTCCACCGGCCCCAGGGGTCACAGAGATCCAGGGAGCCCCCTGGCCCA 531
DB 283 GGAGCCACAGGTGTCCCGGTACAGAGGGCCAGAGGTGATCTTGATTT 342
QY 532 ACTGGCAACAGGACAGAAAGAGAAAGGGGAGCTTGACACCTGGCCCTGGGGT 591
DB 343 CCAGGATGAAGGAGAAAGGTAATTCAGATTTCCAGGACACCTGGACCTCCAGGG 402
QY 592 GAGAGAGGCCCAATTTGACACAGTGTCCCGGAGAGAGCTGGCGGCAAAAGATCTAA 651
DB 403 CAAGTGGACCAAAAGACACCTGATGACGTGAGAGAGCTGGCAGTGAAGATC 462
QY 652 GGCTCCACAGGCCCCCAAGAGCTCCGTGGTCCCTGGGAGAGCCCGCCCTCAGGGCCC 711
DB 463 TCCCTTCAGAGAAAGCCAGGCCACCTGGTTACGTGAGAAACAGGATGACAGAGAA 522
QY 712 AGTGGGAGCCAGGCCCCCGGGCCACAGGCAAGAGAGTCCCGCCCTCAGGGC 771
DB 523 CCGGGGCCCCAGAGACACAGGATCCAGGACCTGTGGGCAAAAGGTAACAGGG 582
QY 772 CCTCTGGCTTCAGGAGACTTCAGGACACCTGGGGAGCTGGGGTGCCTGACCTGG 831
DB 583 GAGGATGTTCACAGCACTCCTGGACCACTGGAGAAAGCAAAAGGTTGTAA 642
QY 832 GGACTCCAGAGCTTGGCTGGGGTACAGGATGCCAGGCCCCAGGGCCCCCGCT 891
DB 643 GGAGAGCAAGAGACACCTGATGATGCTGCTGCCAGGCTTGAAGGGAAACCTGAGAC 702
QY 892 CCTGGCCATCAGAGCGGTGG 913
DB 703 ACTGGACACCTGACAGCAGGG 724

RESULT 13
US-08-392-367B-1

;; Sequence 1, Application US/08392367B
;; Patent No. 5691197
;; GENERAL INFORMATION:
;; APPLICANT: Trygveason, Karl
;; APPLICANT: Elioma, Outi
;; APPLICANT: Kangas, Maarit
;; TITLE OF INVENTION: An Insolated DNA Sequence For a
;; Patent No. 5691197
;; TITLE OF INVENTION: No. 5691197el Macrophage Receptor with
;; TITLE OF INVENTION: a Collagenous Domain and the
;; TITLE OF INVENTION: Polypeptide Chain Encoded by
;; TITLE OF INVENTION: such a Sequence
;; NUMBER OF SEQUENCES: 2
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
;; ADDRESSEE: Minnich & McKee
;; STREET: 1100 Superior Avenue
;; STREET: Suite 700
;; CITY: Cleveland
;; STATE: Ohio
;; COUNTRY: U.S.A.
;; ZIP: 44114-2518
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 inch,
;; MEDIUM TYPE: 720 Kb storable
;; COMPUTER: IBM PS/2, Model 35 SX
;; OPERATING SYSTEM: DOS 5.0
;; SOFTWARE: Word Perfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/392,367B
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Minnich, Richard J.

```

: REGISTRATION NUMBER: 24,175
: REFERENCE/DOCKET NUMBER: TRV 2 009
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (216) 861-5582
: TELEFAX: (216) 241-1666
: TELEX: (216) 980162
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1868 base pairs
: TYPE: Nucleic acid
: STRANDEDNESS: Single
: TOPOLOGY: Linear
: MOLECULE TYPE: Nucleotide-genomic DNA
: HYPOTHEICAL: No. 5691197 relevant
: ANTI-SENSE: No. 5691197 relevant
: US-08-392-367B-1

```

```

Query Match          7.3%; Score 99.4; DB 1; Length 1868;
Best Local Similarity 51.5%; Pred. No. 4.9e-17;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

```

```

QY 469 CAAGGTCCACCGGCCCCCAAGGCTCCGATGACAGAGATCCAGGACCCCTGGC 528
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 CCAGGTGCCAAGAGACCCCAAGGACCCCAAGGACCCCAAGGACCCCTGGC 780
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 529 CCAACTGGCAACAAGGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 588
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 CTTACGGGTGCACAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 589 GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 648
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 GGCAGCAAAAGTGACATAGGTCTCAGTGGCCCAAGGGGGAACATGGCACCAAGGAGAC 900
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 649 AAAGGCTCCCAAGGACCCCAAGGCTCCGATGACAGAGATCCAGGACCCCTGGC 708
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 AAAGGACCTTAGGCTCCCAAGGACCCCAAGGACCCCAAGGACCCCTTAGGCT 960
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 709 CCAAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 768
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 CCAATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 769 GGGCTCTGCTGCTCCAGGAGACTTCAGGACCTTGGGAGCTTGGGAGCTTGGAGCT 828
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 GGTGCTGCTGATCTCCAGGAGTCAAGAGTCAAGAGTCAAGAGTCAAGAGTGT 1080
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 829 CGGAGACTGACAGGCTTCCCTGGGAGTACAGGATGCCAGGACCCCAAGGACCCCTGGC 888
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 CCAAGGCTTCAAGGTGACAGGACTTTCAGAGTCCAGAGGTGAGCCAGGACGACTGT 1140
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 889 CTTCTGCGCCATCAGAGAGCGGTGG 913
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 CTTCTGCGCCAGCAGAGACCCCGG 1165
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 14
US-08-893-467A-1
: Sequence 1, Application US/08893467A
: Patent No. 6063901
: GENERAL INFORMATION:
: APPLICANT: Trygvason, Karl
: APPLICANT: Eliomaa, Outi
: APPLICANT: Kangas, Maarit
: TITLE OF INVENTION: An isolated DNA sequence for a
: Patent No. 6063901
: TITLE OF INVENTION: No. 6063901el Macrophage Receptor with
: TITLE OF INVENTION: a Collagenous Domain and the
: TITLE OF INVENTION: Polypeptide Chain Encoded by
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Fay, Sharpe, Beall, Fagan,
: ADDRESSEE: Minnich & McKee
: STREET: 1100 Superior Avenue

```

```

: STREET: Suite 700
: CITY: Cleveland
: STATE: Ohio
: COUNTRY: U.S.A.
: ZIP: 44114-2518
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch,
: MEDIUM TYPE: 720 Kb storable
: COMPUTER: IBM PS/2, Model 35 SX
: OPERATING SYSTEM: DOS 5.0
: SOFTWARE: Word Perfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/893,467A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Minnich, Richard J.
: REGISTRATION NUMBER: 24,175
: REFERENCE/DOCKET NUMBER: TRV 2 009
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (216) 861-5582
: TELEFAX: (216) 241-1666
: TELEX: (216) 980162
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1868 base pairs
: TYPE: Nucleic acid
: STRANDEDNESS: Single
: TOPOLOGY: Linear
: MOLECULE TYPE: Nucleotide-genomic DNA
: HYPOTHEICAL: No. 6063901 relevant
: ANTI-SENSE: No. 6063901 relevant
: US-08-893-467A-1

```

```

Query Match          7.3%; Score 99.4; DB 3; Length 1868;
Best Local Similarity 51.5%; Pred. No. 4.9e-17;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

```

```

QY 469 CAAGGTCCACCGGCCCCCAAGGCTCCGATGACAGAGATCCAGGACCCCTGGC 528
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 CCAGGTGCCAAGAGACCCCAAGGACCCCAAGGACCCCAAGGACCCCTGGC 780
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 529 CCAACTGGCAACAAGGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 588
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 CTTACGGGTGCACAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 589 GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 648
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 GGCAGCAAAAGTGACATAGGTCTCAGTGGCCCAAGGGGGAACATGGCACCAAGGAGAC 900
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 649 AAAGGCTCCCAAGGACCCCAAGGCTCCGATGACAGAGATCCAGGACCCCTGGC 708
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 AAAGGACCTTAGGCTCCCAAGGACCCCAAGGACCCCAAGGACCCCTTAGGCT 960
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 709 CCAAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 768
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 CCAATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 769 GGGCTCTGCTGCTCCAGGAGACTTCAGGACCTTGGGAGCTTGGGAGCTTGGAGCT 828
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 GGTGCTGCTGATCTCCAGGAGTCAAGAGTCAAGAGTCAAGAGTCAAGAGTGT 1080
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 829 CGGAGACTGACAGGCTTCCCTGGGAGTACAGGATGCCAGGACCCCAAGGACCCCTGGC 888
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 CCAAGGCTTCAAGGTGACAGGACTTTCAGAGTCCAGAGGTGAGCCAGGACGACTGT 1140
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 889 CTTCTGCGCCATCAGAGAGCGGTGG 913
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 CTTCTGCGCCAGCAGAGACCCCGG 1165
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 15
US-08-553-669-11

```

Sequence 11, Application US/08555669
Patent No. 5773248
GENERAL INFORMATION:
APPLICANT: Brewton, Richard G.
APPLICANT: Mayne, Richard
TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/555,669
FILING DATE: 13-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 8389-030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-3660
TELEFAX: 415-854-3694
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2543 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 47..2098
US-08-555-669-11

Query Match 7.3%; Score 99.4; DB 1; Length 2543;
Best Local Similarity 51.5%; Pred. No. 5 6e-17;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 469 CAAGTCCACCGGGCCCAAGGGGTCCAGAGATGACAGAGATCCAGGGACCCCTGGC 528
DB 584 CCAGGTCCCCCAGGGCCCTGGAATGCCAGGGTTCAGAGGACCACTGGCTACAAAGGC 643
QY 529 CCAACTGGCAACAAGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 588
DB 644 GAGCAGGGGGGAAGTCGGGACGAGGACGAGAAAGGTGACCTGGCCCCCTGGGCCGCC 703
QY 589 GGTGAGAGAGGCCCAATTGAGCAGCTGTCCTCCCGAGAGAGAGAGAGAGAGATCA 648
DB 704 GGCTCCCGGGGACGCTGGGGCTGCGAGGGCCCCCGGGGATACGAGAGACTGCCAGGGCCA 763
QY 649 AAAGGCTCCAGGGCCCAAGAGCTCCGTGTTCCCTGGGAAGCCCGGCCCTCAGGGC 708
DB 764 CTCGGGCCCTCGGGGACCGGGGTCCATTTGGGTCCGAGGGCCGCTGGGATCCAGGA 823
QY 709 CCAGTGGGGACCCAGCCCCCGGGGCCACCGAGGCAAGAGAGAGACTCCCGGCCCTCAG 768
DB 824 GCGCTTGGGAAAGCGGGTGACCGAGGCGAGAGGGGCCCAAGAGGTTCCGCGGCCCAAG 883
QY 769 GGGCTCTGAGCTTCAGGAGACTTCAGGGACCGCTGGGAGAGCTGGGGTGCCTGGACCT 828
DB 884 GGTGACCTGGGACGAGCTGTGTCACAGGGAACCCCGGAGTGGCCGGGCCACAGCGAGAG 943
QY 829 CGGGGACTGCCAAGCTTGGCTGGGTACCAAGCATGCCAGGGCCCCCGCCCGGC 888

DB 944 CCGGGCATGCCAGGCAAGGACGCGCCAGAAATGGCTGCCAGAGACTGCATGCCAAGAGGA 1003
QY 889 CCTCTGGGCCATCCAGGAGCGGTGG 913
DB 1004 GAGGCTGTGCGACAGCGTGTCTCGG 1028

Search completed: March 21, 2003, 08:25:41
Job time : 61.326 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:57:56 ; Search time 208.91 Seconds
(without alignments)
5089.177 Million cell updates/sec

Title: US-09-763-712A-1_COPY_325_1695

Perfect score: 1371

Sequence: 1 atgacacgcgcgcacac.....cagracgtcctcgcattca 1371

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEM_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEM_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEM_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEM_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEM_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEM_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEM_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 1369.4 | 99.9 | 2930 | 10 | US-09-745-763-198 |
| 2 | 111 | 8.1 | 6728 | 10 | US-09-954-456-782 |
| 3 | 111 | 8.1 | 6728 | 10 | US-09-880-107-3946 |
| 4 | 110.6 | 8.1 | 1133 | 9 | US-09-924-340-57 |
| 5 | 110.6 | 8.1 | 1133 | 9 | US-09-924-340-57 |
| 6 | 109.4 | 8.0 | 2691 | 10 | US-09-925-302-64 |
| 7 | 107.2 | 7.8 | 2192 | 10 | US-09-925-301-42 |
| 8 | 107 | 7.8 | 5086 | 10 | US-09-880-107-7947 |
| 9 | 107 | 7.8 | 5145 | 9 | US-09-925-299-206 |
| 10 | 107 | 7.8 | 5145 | 10 | US-09-925-299-206 |
| 11 | 107 | 7.8 | 5432 | 12 | US-10-044-090-22 |
| 12 | 106.8 | 7.8 | 4908 | 9 | US-10-001-887-33 |
| 13 | 106 | 7.7 | 2542 | 9 | US-09-954-531-961 |
| 14 | 106 | 7.7 | 2542 | 10 | US-09-954-824A-255 |
| 15 | 105.4 | 7.7 | 5416 | 10 | US-09-954-456-786 |
| 16 | 105.4 | 7.7 | 5416 | 10 | US-09-880-107-2094 |
| 17 | 105.2 | 7.7 | 6158 | 10 | US-09-919-497-6 |
| 18 | 105.2 | 7.7 | 6158 | 10 | US-09-954-456-762 |
| 19 | 104.2 | 7.6 | 4149 | 12 | US-10-044-090-104 |

| | | | | | | |
|----|-------|-----|------|----|--------------------|--------------------|
| 20 | 102.4 | 7.5 | 1486 | 10 | US-09-925-302-247 | Sequence 247, App |
| 21 | 101.6 | 7.4 | 3226 | 10 | US-09-954-456-725 | Sequence 725, App |
| 22 | 101.4 | 7.4 | 3690 | 12 | US-10-044-090-448 | Sequence 448, App |
| 23 | 98.6 | 7.2 | 2158 | 9 | US-10-001-835-98 | Sequence 98, App |
| 24 | 97.8 | 7.1 | 2520 | 10 | US-09-880-107-3685 | Sequence 3685, App |
| 25 | 95.8 | 7.0 | 1485 | 10 | US-09-925-302-246 | Sequence 246, App |
| 26 | 95.2 | 6.9 | 1797 | 9 | US-09-978-295A-613 | Sequence 613, App |
| 27 | 95.2 | 6.9 | 1797 | 9 | US-09-978-295A-613 | Sequence 613, App |
| 28 | 95.2 | 6.9 | 1797 | 9 | US-09-978-192A-613 | Sequence 613, App |
| 29 | 95.2 | 6.9 | 1797 | 9 | US-09-999-832A-613 | Sequence 613, App |
| 30 | 95.2 | 6.9 | 1797 | 9 | US-09-978-189-613 | Sequence 613, App |
| 31 | 95.2 | 6.9 | 1797 | 9 | US-10-174-590-331 | Sequence 331, App |
| 32 | 95.2 | 6.9 | 1797 | 9 | US-10-176-758-331 | Sequence 331, App |
| 33 | 95.2 | 6.9 | 1797 | 9 | US-10-175-737-331 | Sequence 331, App |
| 34 | 95.2 | 6.9 | 1797 | 9 | US-10-173-706-331 | Sequence 331, App |
| 35 | 95.2 | 6.9 | 1797 | 9 | US-10-175-728-331 | Sequence 331, App |
| 36 | 95.2 | 6.9 | 1797 | 9 | US-10-175-752-331 | Sequence 331, App |
| 37 | 95.2 | 6.9 | 1797 | 9 | US-10-176-482-331 | Sequence 331, App |
| 38 | 95.2 | 6.9 | 1797 | 9 | US-10-176-757-331 | Sequence 331, App |
| 39 | 95.2 | 6.9 | 1797 | 9 | US-10-176-913-331 | Sequence 331, App |
| 40 | 95.2 | 6.9 | 1797 | 9 | US-10-180-552-331 | Sequence 331, App |
| 41 | 95.2 | 6.9 | 1797 | 9 | US-10-180-557-331 | Sequence 331, App |
| 42 | 95.2 | 6.9 | 1797 | 9 | US-10-173-700-331 | Sequence 331, App |
| 43 | 95.2 | 6.9 | 1797 | 9 | US-10-174-572-331 | Sequence 331, App |
| 44 | 95.2 | 6.9 | 1797 | 9 | US-10-174-579-331 | Sequence 331, App |
| 45 | 95.2 | 6.9 | 1797 | 9 | US-10-174-582-331 | Sequence 331, App |

ALIGNMENTS

RESULT 1

US-09-745-763-198

Sequence 198, Application US/09745763

Patent No. US2002006394A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

McCoy, John M.

LaVallie, Edward R.

Collins-Racie, Lisa A.

Evans, Cheryl

Merberg, David

Treacy, Maurice

Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

ENCODING THEM

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/745,763

FILING DATE: 18-Jun-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.

REGISTRATION NUMBER: 41,323

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8284

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 198:

SEQUENCE CHARACTERISTICS:

LENGTH: 2930 base pairs

TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-09-745-763-198

Query Match 99.9%; Score 1369.4; DB 10; Length 2930;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 1370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGAACAGCCAGCTCACTATTTCACAGGTCAATGAGAACATCACCATTATCTCCAA 60
DB 921 ATGAACAGCCAGCTCACTATTTCACAGGTCAATGAGAACATCACCATTATCTCCAA 980
QY 61 GCCAAGCAGCAGAACCTGGAAGACCTGACAGCTTACACAAAGTGCAGAGATAGACA 120
DB 981 GCCAAGCAGCAGAACCTGGAAGACCTGACAGCTTACACAAAGTGCAGAGATAGACA 1040
QY 121 GCCATCAAGTTCAACCAACTGAGAGAAAGCTTCCAGCTTTGAGACGATATTGTGAC 180
DB 1041 GCCATCAAGTTCAACCAACTGAGAGAAAGCTTCCAGCTTTGAGACGATATTGTGAC 1100
QY 181 ATGATTGCAATATCAATGTTACACAGCCACCCTGGGAGCGTGCACCAATCTAAAT 240
DB 1101 ATGATTGCAATATCAATGTTACACAGCCACCCTGGGAGCGTGCACCAATCTAAAT 1160
QY 241 GAAGTCAGAGCACTTGCACAGATACCTTACCAAAACACAGATGATGACCTCCTTG 300
DB 1161 GAAGTCAGAGCACTTGCACAGATACCTTACCAAAACACAGATGATGACCTCCTTG 1220
QY 301 AATAATACCTGGCCAAACATCCGTTGGATTCTGTTTCTCTCAAGATGACAAAGATTGG 360
DB 1221 AATAATACCTGGCCAAACATCCGTTGGATTCTGTTTCTCTCAAGATGACAAAGATTGG 1280
QY 361 ATGAGTCGAGGTTAGACATGAAAGTGAAGTCACTTATCATGATTAAGAAATGAAG 420
DB 1281 ATGAGTCGAGGTTAGACATGAAAGTGAAGTCACTTATCATGATTAAGAAATGAAG 1340
QY 421 CTAGTAGACTCCAAAGATGCTGACGCTCATCAGAAATTTTCAATACTACAGAGTCCAC 480
DB 1341 CTAGTAGACTCCAAAGATGCTGACGCTCATCAGAAATTTTCAATACTACAGAGTCCAC 1400
QY 481 GGGCCCCAGGGGTCCAAAGAGTGCACAGAGATCCAGGGACCCCTGGCCCACTGGCAAC 540
DB 1401 GGGCCCCAGGGGTCCAAAGAGTGCACAGAGATCCAGGGACCCCTGGCCCACTGGCAAC 1460
QY 541 AAGGACAGAAAGAGAGAGAGGGGAGGACCTGAGACCACTGGCCCTGGGGTGAAGAGGC 600
DB 1461 AAGGACAGAAAGAGAGAGAGGGGAGGACCTGAGACCACTGGCCCTGGGGTGAAGAGGC 1520
QY 601 CCAATTGGACCACTGCTGCCCGGAGAGAGCTGGGCGGCAAAAGATCTAAAGGCTCCAG 660
DB 1521 CCAATTGGACCACTGCTGCCCGGAGAGAGCTGGGCGGCAAAAGATCTAAAGGCTCCAG 1580
QY 661 GGGCCCAAAGGCTCCGTTGCTCCCTGGGAGAGCCCGGCTTACAGGGCCCAAGTGGGAC 720
DB 1581 GGGCCCAAAGGCTCCGTTGCTCCCTGGGAGAGCCCGGCTTACAGGGCCCAAGTGGGAC 1640
QY 721 CCAAGGCCCCCGGGGACCAAGAGAGAGGACTCCCGGCTTCAAGGGCCCTCTCTGGGC 780
DB 1641 CCAAGGCCCCCGGGGACCAAGAGAGAGGACTCCCGGCTTCAAGGGCCCTCTCTGGGC 1700
QY 781 TTCAGGAGCTTCAAGGACACCTGTTGGGAGAGCTGGGAGCTTGAACCTTCGGGAGTGC 840
DB 1701 TTCAGGAGCTTCAAGGAGACCTGTTGGGAGAGCTGGGAGCTTGAACCTTCGGGAGTGC 1760
QY 841 GGGCTTGGCTGGGGTACAGGAGATGCCAGGCCCAAGGGCCCGCCCTCTCTGGGCCA 900
DB 1761 GGGCTTGGCTGGGGTACAGGAGATGCCAGGCCCAAGGGCCCGCCCTCTCTGGGCCA 1820
QY 901 TCAGGAGCGGTGGTCCCGCTGGGCTGCAGATGAGCAACCCCGGAGCGGAGAGCAAT 960
DB 1821 TCAGGAGCGGTGGTCCCGCTGGGCTGCAGATGAGCAACCCCGGAGCGGAGAGCAAT 1880
```

```
QY 961 GGCTGCCCGCTCACTGGAAGACTTCACAGACAAATGCTACTATTTTCAGTTGAGAAA 1020
DB 1881 AGCTGCCCGCTCACTGGAAGACTTCACAGACAAATGCTACTATTTTCAGTTGAGAAA 1940
QY 1021 GAAATTTTGGAGATCAAGGCTTTTCTGTGTGAAGACAGTCTTCACATCTTGTTCATA 1080
DB 1941 GAAATTTTGGAGATCAAGGCTTTTCTGTGTGAAGACAGTCTTCACATCTTGTTCATA 2000
QY 1081 AACACTAGAGAGGACAGCAATGATAAAAACATAGGTAGGAGAGAGAGCCACTGG 1140
DB 2001 AACACTAGAGAGGACAGCAATGATAAAAACATAGGTAGGAGAGAGAGCCACTGG 2060
QY 1141 ATCGGCTTCACAGACTCAGAGCGTGAATAATGAAATGGAAGTGGCTGGATGGACATCTCA 1200
DB 2061 ATCGGCTTCACAGACTCAGAGCGTGAATAATGAAATGGAAGTGGATGGACATCTCA 2120
QY 1201 GACTACAAAATTTGAAAGCTGACAGCGGATTAAGTGGGCTATGCGCATGGGCCAGGA 1260
DB 2121 GACTACAAAATTTGAAAGCTGACAGCGGATTAAGTGGGCTATGCGCATGGGCCAGGA 2180
QY 1261 GAAGACTGTGCTGGTTGATTATGCTGGGCGAGTGGAGAGATTCATATGTGAAGAGCTC 1320
DB 2181 GAAGACTGTGCTGGTTGATTATGCTGGGCGAGTGGAGAGATTCATATGTGAAGAGCTC 2240
QY 1321 AATACTTCATTTGCGAAAAGACAGGAGACAGTACTGATCATCTGCATTA 1371
DB 2241 AATACTTCATTTGCGAAAAGACAGGAGACAGTACTGATCATCTGCATTA 2291
```

RESULT 2

US-09-954-456-782
Sequence 782, Application US/09954456
Patent No. US20020115057A1

GENERAL INFORMATION:

APPLICANT: Young, Paul

TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C

FILE REFERENCE: 689290-76

CURRENT APPLICATION NUMBER: US/09/954,456

PRIOR FILING DATE: 2001-09-18

PRIOR APPLICATION NUMBER: US/60/233,617

PRIOR FILING DATE: 2000-09-18

PRIOR APPLICATION NUMBER: US/60/234,052

PRIOR FILING DATE: 2000-09-20

PRIOR APPLICATION NUMBER: US/60/234,923

PRIOR FILING DATE: 2000-09-25

PRIOR APPLICATION NUMBER: US/60/235,134

PRIOR FILING DATE: 2000-09-25

PRIOR APPLICATION NUMBER: US/60/235,637

PRIOR FILING DATE: 2000-09-26

PRIOR APPLICATION NUMBER: US/60/235,638

PRIOR FILING DATE: 2000-09-26

PRIOR APPLICATION NUMBER: US/60/235,711

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: US/60/235,720

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: US/60/235,840

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: US/60/235,863

PRIOR FILING DATE: 2000-09-27

NUMBER OF SEQ ID NOS: 2276

SOFTWARE: PatentIn version 3.0

SEQ ID NO 782

LENGTH: 6728

TYPE: DNA

ORGANISM: Homo sapiens

US-09-954-456-782

Query Match 8.1%; Score 111; DB 10; Length 6728;

Best Local Similarity 53.0%; Pred. No. 2,6e-20;
Matches 237; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

| | | | |
|----|------|---|------|
| OY | 467 | TACAACTCCACCGGGGCCCAAGGATCACAAGATCCACAGGATCCCGCTG | 526 |
| Db | 721 | TCGAAGTCCCCCTGGTGACCCTCGGCAGCCTTGACTTCAGTGTCCCAG | 780 |
| OY | 527 | GCCCACTGCACACAGGACAGAAGAGAGAGAGGAGGAGCCTGACCTGGCCCTG | 586 |
| Db | 781 | GTCCCCCAGGTGCCCTCGAATAAGATGATGATGAGAACCTGGAAACCTGTGCTVC | 840 |
| OY | 587 | CGGTTGAGAGGGCCCAATTGAGCACAGCTGATGCCCGGACAGAGTGTCGGCAAGAT | 646 |
| Db | 841 | CTGCTGAGCCTGGGCTCCTCGGSCCTCAGGGCTCTGAGAGATTGCCCGAACAGTGGCC | 900 |
| OY | 647 | CTAAAGGCTCCACAGGGGCCCAAAGGCTCCCTGCTTCCCTGGAGAGCCCGCCTCAG | 706 |
| Db | 901 | TCCCTGAAATGAAGGACACAGAGGTTTCAGTGTTGATGGTCCAAAGSAGATGCTG | 960 |
| OY | 707 | GCCCCATGGGGAGCCACAGGCCCCCGGGGCCACACAGGACAAAGAGGACTCCCGGCTTC | 766 |
| Db | 961 | GTCTCTGTTGCTTAAGAGGTGAGCCCTGGAGCCCTGTGTAAAATGGACTCTCTGTCTAGA | 1020 |
| OY | 767 | AGGCGCTCTGTGGCTTCACAGGACTTCACAGGACACGCTTGGGAGGCTGGGGTCTGTGAC | 826 |
| Db | 1021 | TGGGGCCCCCGGTGGCTCCCTGGTGAAGAGAGTGCCTCCAGACCCCTGTGGCTGTG | 1080 |
| OY | 827 | CTGGGAGCTGCAGAGCTTGCCTGGGGGTACCAAGCATGCTCCAGGCCCCCAAGGGCCCCCG | 886 |
| Db | 1081 | CTCTGGAAATGATGTGCTACTGCTGCTGCGGGCCCCCTGTGCCACGGCCCCGCTG | 1140 |
| OY | 887 | GCCCTCTGGGCCATCAGAGCGGTGG | 913 |
| Db | 1141 | GTCTCTGGGCTTCCCTGGTGGCTGGTGG | 1167 |

[illegible]

| | | | |
|----|------|---|------|
| Db | 841 | CTGGTAGACGCTGGGGCCCTCTCGGGCCCTCAGGGTCTCTGAGAGATTGGCCCGGAAACAGCTGGCC | 900 |
| Qy | 647 | CTAAAGGCTCCCAAGGCCCCAAGAGGCTCCCGTGGTTCCTCCCTGGGAAGCCCGGCTCTAGG | 706 |
| Db | 901 | TCCCTGGAATGAAGAGGACACAGAGAGTTTCAGTGGTTGGATGGTCCCAAGGGAGATGCTG | 960 |
| Qy | 707 | GCCCCAGTGGGGAGCCAGGGCCCCCGGGCCACACAGGCAAGAGGAGATCCCGGGCCCTC | 766 |
| Db | 961 | GTTCTGCTGTTCTTAAGGGTGAAGCTGGCAGCCCTGTGTAAATATGAGATCTCTGTGCAGA | 1020 |
| Qy | 767 | AGGGCCCTCTGAGCTCCAGGAGATTCAAGGGCACCGTTGGGAGAGCCTGGGGTGCCTGGAC | 826 |
| Db | 1021 | TGGGGCCCCGTTGGCTCTGTGTAGAGAGAGTGCAGCCCTGAGACCCCTGGCCCTGTGTGGTG | 1080 |
| Qy | 827 | CTCGGGGACATCCAGGCTTGTGCTTGGGGTACCAAGGATGCGACAGGCCCCCAAGAGGGCCCCCGG | 886 |
| Db | 1081 | CTCTGTGAAATGATGTGCTACTACGTGTGCGCGGGCCCCCTGTGCCACACGGCCCCCGCTG | 1140 |
| Qy | 887 | GACCTCTGGGCCATCAGAGAGCGGTGG | 913 |
| Db | 1141 | GTCTCCTCGGCTTCCCTGGTGTGCTGG | 1167 |

| | | | |
|-----------------------|------|--|-----|
| Db | 169 | CAGAGGAGAGCGCAGAGACCCCAAGGTACTCGGCTGAGAAAGGCGACACAGGATTTTCAGAG | 228 |
| QY | 495 | AGAGAGGTACACAGAGATCCACAGGACCCCTGTGCCCAACTGGCAACAGACAGAAAGG | 554 |
| Db | 109 | GCMAAGAGCGAGGCCCCCAGGATCTGAAGGCTCCAGGCCCCCCGAGCCGAGCGTCC | 168 |
| QY | 435 | GCATGCTACGCTCATCAGATTTTACATTACTACTACAGSTCCACCGGGCCCCAGGGGTCC | 494 |
| Query Match | | 8.1%; Score 110.6; DB 9; Length 1133; | |
| Best Local Similarity | | 52.9%; Pred. NO. 1.2e-20; | |
| Matches | 267; | Conservative 0; Mismatches 229; Indels 9; Gaps 1. | |

```

OY 555 AGAAGAGGGGAGCTGGACCACTGGCCCTGGCGGTGAGAGAGGCCCAATTGGACGAC 614
    || || || || || || || || || || || || || || || || || || || || ||
Db 229 CCAAGCCAGGCTTTCGCGGCCACCGGGTCCCTGGATTCACAGCAAGTTGGATCACC 288
OY 615 TGTGTCCTCCCGG-----AGAGCTGGCGGCAAAAGATCTAAAGGCTCCCGAGGCC 665
    || || || || || || || || || || || || || || || || || || || || ||
Db 289 TGGCCCACTGGCTCCTCAAGCAGAGAGGAGGCAAGGATTCGAGGCCATCAGGCT 348
OY 666 CAAAGCTCTCCGTGTTCCCTGGAGAGCCGCGCTCAGAGGCCCGAGTGGAGCCAGG 725
    || || || || || || || || || || || || || || || || || || || || ||
Db 349 GCCTGGTCCCTCTGGGACCGGACCTCTGGGATTCAGAGGCCCGCGCTCGATGG 408
OY 726 CCCCCCGGCGCCACAGCAAGAGGAGCTCCCGGCTCAGAGGCCCTCTGGCTTCA 785
    || || || || || || || || || || || || || || || || || || || || ||
Db 409 TTTGATGAGGAAGATGCAAGCCTGGCTTGAAGGGGAGACCTGTCTGCTGGCCCCC 468
OY 786 GGAAGCTCAGGCGACCTTGGGAGAGCTGGGCTGACCTGGGAGCTGGGAGCT 845
    || || || || || || || || || || || || || || || || || || || || ||
Db 469 TGGACTATGGGACACCGGCTTTAAGGGGAAACAGACATCTGCTGCCACGAGACC 528
OY 846 GCTGAGGCTACAGGCAATGCGAGGCGCCCGGCGCTCTGAGCCATCAGG 905
    || || || || || || || || || || || || || || || || || || || || ||
Db 529 TAAGGGTACTGTGCAACACAGGTCTCTCTGGCAGCAGTGGCGGCTGGCGAGAGG 588
OY 906 AGCGGTGTGCTCCCTGGCTGAG 930
    || || || || || || || || || || || || || || || || || || || || ||
Db 589 TGAACCTGTGTCATGGAGACCCAG 613
    || || || || || || || || || || || || || || || || || || || || ||

```

RESULT 5 US-09-992-600A-57

Sequence 57, Application US/0992600A
Publication No. US2003027161A1

GENERAL INFORMATION:

APPLICANT: Benjatin, Stephane
APPLICANT: Tanaka, Hiroaki
TITLE OF INVENTION: HUMAN CDMS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91.USA. DIV
CURRENT APPLICATION NUMBER: US/09/992,600A
CURRENT FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: PCT/IB01/01715
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 114
SOFTWARE: JPatent
SEQ ID NO 57
LENGTH: 1133
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..475
FEATURE:
NAME/KEY: CDS
LOCATION: 476..964
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 965..1133
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 1101..1106
FEATURE:
NAME/KEY: polyA_site

```

; LOCATION: 1118..1133
US-09-992-600A-57
Query Match      8.1%; Score 110.6; DB 9; Length 1133;
Best Local Similarity 52.9%; Pred. No. 1,2e-20;
Matches 267; Conservative 0; Mismatches 229; Indels 9; Gaps 1;

```

```

OY 435 GCATGGTCACTCATCAAGATTTTACAAATACAGATCAACGAGGCCCGAGGGGTCC 494
    || || || || || || || || || || || || || || || || || || || || ||
Db 109 CAAAGACAGAGGCGCCCGCCAGATCTGAAAGGCTCCAGGCGCCCGCCAGCGGGTCC 168
OY 495 AAGAGGTGACAGAGATCCCAAGGACCCCTGGCCCACTGGCAACAAGGACAGAAAG 554
    || || || || || || || || || || || || || || || || || || || || ||
Db 169 CAGAGGAGACAGAGAGCCCAAGGTAACTCGGTGAGAAAGGGGACCGAGGATTTCAAG 228
OY 555 AGAAGAGGGGAGAGCTGAGACCTGAGCCCTGGGAGTGAAGAGAGGCCCAATTGGAGCAG 614
    || || || || || || || || || || || || || || || || || || || || ||
Db 229 CCAAGCCAGGCTTTCGCGGCCACCGGGTCCCTGGATTCACAGCAAGTTGGATCACC 288
OY 615 TGTGTCCTCCCGG-----AGAGCTGGCGGCAAAAGATCTAAAGGCTCCCGAGGCC 665
    || || || || || || || || || || || || || || || || || || || || ||
Db 289 TGGCCCACTGGCTCCTCAAGCAGAGAGGAGGCAAGGATTCGAGGCCATCAGGCT 348
OY 726 CCCCCCGGCGCCACAGCAAGAGGAGCTCCCGGCTCAGAGGCCCTCTGAGCTTCA 785
    || || || || || || || || || || || || || || || || || || || || ||
Db 409 TTTGATGAGGAAGATGCAAGCCTGGCTTGAAGGGGAGACCTGTCTGCTGGCCCCC 468
OY 786 GGAAGCTCAGGCGACCTTGGGAGAGCTGGGCTGACCTGGGAGCTGGGAGCT 845
    || || || || || || || || || || || || || || || || || || || || ||
Db 469 TGGACTATGGGACACCGGCTTTAAGGGGAAACAGACATCTGCTGCCACGAGACC 528
OY 846 GCTGAGGCTACAGGCAATGCGAGGCGCCCGGCGCTCTGAGCCATCAGG 905
    || || || || || || || || || || || || || || || || || || || || ||
Db 529 TAAGGGTACTGTGCAACACAGGTCTCTCTGGCAGCAGTGGCGGCTGGCGAGAGG 588
OY 906 AGCGGTGTGCTCCCTGGCTGAG 930
    || || || || || || || || || || || || || || || || || || || || ||
Db 589 TGAACCTGTGTCATGGAGACCCAG 613
    || || || || || || || || || || || || || || || || || || || || ||

```

RESULT 6

US-09-925-302-64
Sequence 64, Application US/09925302
Patent No. US2002004491A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 64
LENGTH: 2691
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (2653)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (2667)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (2683)

| | | | |
|----|------|--|------|
| Db | 829 | AGCTGGTGCCTGGCTGGCTAGTGATGGAAGTGGAGTGGCCCTGGGTCCCTGGCTGGTCCCATTTGG | 888 |
| QY | 645 | ATCTAAGAGGCTCCGAGGGCCCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCA | 704 |
| Db | 889 | GTCCTGGGCCCTCCAGGCTTCCAGATGGCCCTGGCCCCCAAGGTGAAATTTGGAGCTGT | 948 |
| QY | 705 | GGGGCCCCAGTGGGGACCCAGAGCCCCCGGGGCCACAGGCAAGAGAGGACTCCCGGGCC | 764 |
| Db | 949 | TGGTAACGCTGGTCTCTGTGTCCCGCGGTCCCGCTGGTGAAGTGGGCTCTCCAGGGCT | 1008 |
| QY | 765 | TCAGAGGCCCTCTCGCTTCCAGGACTTCAGGSCACCGTTGGGGAGACCTGGGGTCCGTGG | 824 |
| Db | 1009 | CTCCGGGCCCGTTGGACTCTCTGTATCTCTGAGCAAAAGGACCTTACGTGGCCAAAGG | 1068 |
| QY | 825 | ACCTCGGGAGACTGTCAGGCTTGGCTGGGGTACCAAGGACATGCAGAGGCCCAAGGGCCCGC | 884 |
| Db | 1069 | TGCTGTCTGGCTTCCGGGCTTGTCTGGGGCTCCCGGCTCTCTTGGAACCCCGGATTTCC | 1128 |
| QY | 885 | CGGCGCTCTGGCCCATCAGAGAGCGGTGTGTGCCCTCGGCCCTGCAGATAGACCAACCC | 944 |
| Db | 1129 | TGGCCCTGTTGGTGTCTCCGGTCTACTGGTGCCAGAGACACTTTGTGGAGACCTGCTGCC | 1188 |
| QY | 945 | GGC 947 | |
| Db | 1189 | AGC 1191 | |

```

US-09-925-299-206
: RESULT 9
: Sequence 206, Application US/09925299
: Publication No. US2003040617A9
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: P102
: CURRENT APPLICATION NUMBER: US/09/925,299
: CURRENT FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05883
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 1556
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 206
: LENGTH: 5145
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (4)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (17)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (5126)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (5143)
: OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

```

| | | | | |
|-----------------------|--------------|--------------------|-----------------|--------------|
| Query Match | 7.8%; | Score 107; | DB 9; | Length 5145; |
| Best Local Similarity | 51.3%; | Pred. No. 2.9e-19; | | |
| Matches 248; | Conservative | 0; | Mismatches 235; | Indels 0; |
| | | | Gaps | 0; |

| | | |
|----|---|----|
| Oy | ACTACAGGTCCACCGGGCCCCAAGGTCCTCAAGAGTGACAGAAGATCCAGAGACC | 52 |
| | " " " " " " " " " " " " " " " " | |
| Dd | ACAGCCCCGTCCTCCTGGTGTGAAGGGGTAACCTGTGCCCTTGTTAAATGTAACCTC | 77 |
| | " " " " " " " " " " " " " " " " | |
| Oy | TGGCCCAACTGCACAAGGACACAGAAGGAGGAGGAGCTTGACACCACTGGCCC | 58 |
| | " " " " " " " " " " " " " " " " | |

| | | | |
|----|------|---|------|
| Db | 772 | AGGTCAACAGAGACCCGTGGGGCTTCTGGTGAAGAGACGTTGGTGGCCCTGGGCC | 831 |
| QY | 585 | TGCGGTTGAGAGAGAGCCCAATTGGACAGCTGTGTCCTCCCGAGAGACCTGGCGGCAAGG | 644 |
| Db | 832 | AGCTGGTCCCGCTGGCGAGTATGGAGTATGGGGTCCCGTGGGGCTCTGCTGGTCCATTGG | 891 |
| QY | 645 | ATCTAAAGAGTCCCAAGAGGCCCAAGAGCTCCGTGGTTCCTCCCGGGAAGCCCGGCCCTCA | 704 |
| Db | 892 | GTCGTGTGGCCCTCCAGAGCTTCCCAAGTGGCCCTGGGCCCAAGAGGTGAATTTGAGACTGT | 951 |
| QY | 705 | GAGGCCCAGTGGGAGACCAAGGCCCTCCCGGCCCAACAGGCAAGAGAGGACTCCCGGCC | 764 |
| Db | 952 | TGTTAAAGCTGTGTCGTCTGTGGTCCCGCGGTCCCTCGGTGAAGTGGGTCTTCCAGGCT | 1011 |
| QY | 765 | TCAGGGCCCTCTGGCTTCCAGGAGACTTCAGGGACACCGTTGGGGACCTGGGGTGTGCTGG | 824 |
| Db | 1012 | CTCGGGCCCGTGTGACCTCTCTGTAATCTTGAGCAAAAGGCGCTTACGTGGTCCAAAGG | 1071 |
| QY | 825 | ACCTTCGGGAGATGTGCAGAGCTTGGCTGGGGATACAGGACATGCACGAGCCCAAGGGCCGCC | 884 |
| Db | 1072 | TGCTGTGGGCTTCCGGCGTGTGTGGGGCTCCCGGCTCTCCGAGACCCCGGGATATTC | 1133 |
| QY | 885 | CGGCGCTCTGGGCCATTCAGAGAGCGGTGTGCCCTTGCCCTTCGACAGATGAGCCAAAGCC | 944 |
| Db | 1132 | TGGCGCTGTGGTGTCTCCGGTGCTACTAGTGCACAGAGACTGTGTGTGAGACCTGTGTCC | 1191 |
| QY | 945 | GGC 947 | |
| Db | 1192 | AGC 1194 | |

```

RESULT 10
US-09-925-299-206
; Sequence 206, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2001-08-10
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

```

| | | | | |
|-----------------------|--------------|--------------------|-----------------|--------------|
| Query Match | 7.8%; | Score 107; | DB 10; | Length 5145; |
| Best Local Similarity | 51.3%; | Pred. No. 2.9e-19; | | |
| Matches 248; | Conservative | 0; | Mismatches 235; | Indels 0; |
| | | | Gaps | 0; |

| | | | |
|----|-----|---|-----|
| Qy | 465 | ACTCAAGTCCACCGGGCCCCGAGGTCCAAAGTGACAGAGATCCCAGGACC | 524 |
| | | | |
| Db | 712 | ACAGCCCGTGCTCCTGTTGAAGGGTGAACCTGTCGCCCTGTGAAATGGACTCC | 771 |

| QY | 771 | CCCTCCCTGGCTTCCACAGGACCTTACAGGGCACCCTTGGGGAG-----CCTGGGGTCTTGG | 824 |
|----|------|---|------|
| Db | 1903 | CCTGAAGGGGGAACACAGGCTCTCCACAGGGCCCCCTTGAAGGGGAGAGACGAGGGGAACCTGG | 1962 |
| QY | 825 | ACCTCGGGGACTGCCAGGCTTGCCTGGGGGTACCAAGCATGCCAGGCCCCCAAGAGCCCCC | 884 |
| Db | 1963 | CACGGCTTGGGGCCACAGGGGGCCCCCAGAGGGGTCCCTGGCTCCCTGTGAATCACGGGCTTCC | 2022 |
| QY | 885 | CGGCGCTTCTGGCCCATAGAGAGCGGTGTGCCCCCTGGGCCCTCGCAGAAATGAGCCAACCCC | 944 |
| Db | 2023 | GGGGGCTTCCGGGGCCCCCGGGAGCCCTGTGTGCTGCCCTTGGGGGCTTTCGATGAGACTGGCAT | 2082 |
| QY | 945 | GGCA 948 | |
| Db | 2083 | CGCA 2086 | |

| QY | 467 | TACAAGGTCCACCGGGCCCCCAGGGTCCACAGAGGTGACAGAGGATCCACAGGACCCCCCTG | 526 |
|----|-----|--|-----|
| Db | 630 | TCCTCGACACACCGGGTCCACAGGGCATCAGGGGTACCCAGGCAATGACAGGGGCCCAAGG | 689 |
| QY | 527 | GCCCACTGTGCAACAGGAGCAAGAAAGAGAGAGGGGGAGCCTTGACCACTGTGCCCTG | 586 |
| Db | 690 | GAGAGACGGGCTCATGTATATAAGGCATGTGTGGCGCTATCGTCCACATCGAGCCAC | 749 |
| QY | 587 | CGGGGAGACAGAGGCCCATTTGGACCACTGTGCTCCCCCGGAGAGAGCTGTGGCGCAAGAT | 646 |
| Db | 750 | CGGGGAGAGAGGTCTTATGGGGAGCCGACCGCAAGCTGTGGGAGAAAGGGTACGAGGGCA | 809 |
| QY | 647 | CTAAAGGTCCACAGGGCCCCCAAGGCTCCGTTGTTCCCTTGGGAAGCCGAGCCCTCAGG | 706 |
| Db | 810 | GCCAGGATATTCGTGTGACCCACAGGGGATCACAGGCCCAAGAGAGCAACAGGGGCCCCACG | 869 |
| QY | 707 | GCCCCAGTGGGAGCCAGGCCCCCGGGGCCACCAAGGCAAGAGAGGACTCCCGGCTC | 766 |
| Db | 870 | GCATCAACAGGCAAGATGGAGCCCAAGGACAGCCCTGTGCATGAAGGGCAGTGCAGAGCAGG | 929 |
| QY | 767 | AGGGCCCTCTGCGCTTCAGGGACTTCAGGGCACCCTGTGGGAGAGCGGTGCTGTGAC | 826 |

| Query Match | Best Local Similarity | Score | DB | Length |
|---|-----------------------|-------|-----|--------|
| Matches 236; Conservative 0; Mismatches 217; Indels 0; Gaps 0; | 52.1%; | 106; | 10; | 2542; |
| | | | | |
| 467 TACAGGTCACCGGCGCCACAGAGGTCCAGAGGTGACAGAGGTCCAGGACCCCTG | 526 | | | |
| 630 TCCCGAGACACCGGGTCCCGAGGACATCAGGGCTACCCAGGATGGAGGCCCAAG | 689 | | | |
| 527 GCCCACTGCGACAGGAGCAGAAAGAGAGAGGGGAGCCTGGACACCTGGCCCTG | 586 | | | |
| 690 GAGAGACGGCGCTCATGATATTAAGGATGAGGGGCTGATGCTGCTGCTGCTG | 749 | | | |
| 587 CGGGTGAAGAGGCGCATATGACAGCTGTGCTCCCGCGAGAGCGTGGGCGCAAGAT | 646 | | | |
| 750 CGGGTGAAGAGGCTCTTAGGGGAGCGCCAGGCTCGAGCTGGGGAGAAAGGTATACAGG | 809 | | | |
| 647 CTAAAGCTCCAGGCGCCCAAGAGCTCCGCTGCTCCCTGGGAAAGCCCGCCCTGAG | 706 | | | |
| 810 GCCCAGGATTCGTGAGACCCCGAGGAGATCACAGGCGCCGAAAGAGAGCAAGGGCCCGAG | 869 | | | |
| 707 GCCCAGTGGGAGACCCAGGCGCCCGGGGCGCCACAGGAGTAAAGAGGACTCCCGGCGCTC | 766 | | | |
| 870 GCATCAAGCGCAAGAGTGGAGCCCGAGCGCTGTGGCATAAAGGAGATGACGAGACAG | 929 | | | |
| 767 AGGCGCTCTCTGCGCTTCAGAGGACTTCAGGGCACCGTTGGGGAGACTTGGGGTGGTGGAC | 826 | | | |
| 930 CGGGACACCGCGGAAGTCCAGGACACAGGGCTCTAGGGGCTGTGCTCAGGCGCCTGGGA | 989 | | | |
| 827 CTCGGGAGCTCCAGGCTTGGCTTGGGATACAGGAGATGCCAGGCCCAAGGGCCCCCGG | 886 | | | |
| 990 CAAAGAGGCGCTGGAGAACCGAGGTGAGCCGGGCGCAGAGGCTTCTGTGATTTCTGTG | 1049 | | | |
| 887 GCCCTCTGCGCCCATCAGAGCGGTGTGCCCC | 919 | | | |
| 1050 GTCCCCCTGGGAAGAGGAGAGCCAGGCGCTC | 1082 | | | |

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:34:06 ; Search time 1798.64 Seconds
(without alignments)
12344.923 Million cell updates/sec

Title: US-09-763-712A-1_COPY_325_1695

Perfect score: 1371

Sequence: 1 atgacagccagctcaactc.....cagctactgtcatctgcatcta 1371

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estlum:*
3: em_estlin:*
4: em_estlmu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estlro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 1084.6 | 79.1 | 3305 | 11 | BC009162 Mus muscu |
| 2 | 677.8 | 49.4 | 957 | 14 | B0955927 AGENCOURT |
| 3 | 581.2 | 42.4 | 936 | 14 | B0891432 AGENCOURT |
| 4 | 559 | 40.8 | 906 | 13 | B1456109 AGENCOURT |
| 5 | 556.4 | 40.6 | 861 | 14 | B0713873 AGENCOURT |
| 6 | 525.4 | 38.3 | 668 | 14 | B0674807 AGENCOURT |

| RESULT 1 | LOCUS | DEFINITION | ACCESSION | VERSION | KEYWORDS | SOURCE | ORGANISM | REFERENCE | AUTHORS | TITLE | JOURNAL | REMARK | COMMENT |
|----------|----------|--|-----------|------------|-------------|--------|--------------|---------------------|---------------|------------------|--|---|---------|
| BC009162 | BC009162 | Mus musculus, Similar to scavenger receptor with C-type lectin, clone IMAGE:2811487, mRNA. | BC009162 | BC009162.1 | GI:14714370 | HTC. | house mouse. | 1 (bases 1 to 3305) | Strausberg,R. | Direct Submision | Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA | NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ | |

ALIGNMENTS

| C | 7 | 475.4 | 34.7 | 967 | 9 | AL568743 |
|----|-------|-------|------|-----|-----------|-----------|
| 8 | 445.6 | 32.5 | 580 | 12 | BE910803 | AL568743 |
| 9 | 437.2 | 31.9 | 552 | 10 | AW958053 | BE910803 |
| 10 | 415.8 | 30.3 | 601 | 10 | BE290299 | BE910803 |
| 11 | 385.6 | 28.1 | 884 | 14 | B0934501 | B0934501 |
| 12 | 369.4 | 26.9 | 638 | 10 | BE248064 | BE248064 |
| 13 | 351.4 | 25.6 | 507 | 14 | BM675508 | BM675508 |
| 14 | 350 | 25.5 | 507 | 14 | BM713891 | BM713891 |
| 15 | 343.8 | 25.1 | 808 | 14 | B0771366 | B0771366 |
| 16 | 311 | 22.7 | 339 | 9 | AA304251 | AA304251 |
| 17 | 301.4 | 22.0 | 357 | 10 | AW379436 | AW379436 |
| 18 | 297.2 | 21.7 | 517 | 9 | AL543000 | AL543000 |
| 19 | 296 | 21.6 | 310 | 9 | AA361740 | AA361740 |
| 20 | 295.2 | 21.5 | 679 | 14 | W27610 | W27610 |
| 21 | 281.2 | 20.5 | 581 | 14 | B0127513 | B0127513 |
| 22 | 271 | 19.8 | 350 | 9 | AA012704 | AA012704 |
| 23 | 258.2 | 18.8 | 286 | 12 | BE854324 | BE854324 |
| 24 | 245.8 | 17.9 | 420 | 12 | BE929813 | BE929813 |
| 25 | 244 | 17.8 | 277 | 12 | BE854320 | BE854320 |
| 26 | 233.8 | 17.1 | 429 | 12 | BE929798 | BE929798 |
| 27 | 230.8 | 16.8 | 465 | 10 | AM240221 | AM240221 |
| 28 | 223.2 | 16.3 | 280 | 13 | B01010883 | B01010883 |
| 29 | 212.6 | 15.5 | 281 | 12 | BE854408 | BE854408 |
| 30 | 212 | 15.5 | 723 | 14 | B0004176 | B0004176 |
| 31 | 210.2 | 15.3 | 943 | 14 | B0887163 | B0887163 |
| 32 | 205.8 | 15.0 | 261 | 13 | B1010880 | B1010880 |
| 33 | 205.6 | 15.0 | 1082 | 12 | BM907108 | BM907108 |
| 34 | 201 | 14.7 | 655 | 12 | BG482931 | BG482931 |
| 35 | 200.8 | 14.6 | 345 | 9 | A1742661 | A1742661 |
| 36 | 200.8 | 14.6 | 410 | 12 | BE856923 | BE856923 |
| 37 | 199.6 | 14.6 | 441 | 10 | BE749743 | BE749743 |
| 38 | 193.4 | 14.1 | 234 | 12 | BE854188 | BE854188 |
| 39 | 192.2 | 14.0 | 511 | 12 | BE805750 | BE805750 |
| 40 | 185.2 | 13.5 | 591 | 13 | B1445884 | B1445884 |
| 41 | 182.2 | 13.3 | 839 | 13 | B1697412 | B1697412 |
| 42 | 175 | 12.8 | 201 | 14 | BQ367776 | BQ367776 |
| 43 | 172.4 | 12.6 | 334 | 10 | AW002141 | AW002141 |
| 44 | 170.4 | 12.4 | 309 | 14 | W76214 | W76214 |
| 45 | 144.8 | 10.6 | 592 | 12 | BG175047 | BG175047 |

Contact: amg@bcm.tmc.edu
Gunnarntre, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richard, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRK Plate: 5 Row: h Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis

This clone has the following problem: incomplete processing.

FEATURES

source
Location/Qualifiers
1..3305
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="IMAGE:2811487"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI_CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
BASE COUNT 1031 a 757 c 760 g 757 t
ORIGIN

Query Match 79.1%; Score 1084.6; DB 11; Length 3305;
Best local Similarity 86.9%; Pred. No. 9.5e-257;

Matches 1192; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 1 ATGACACGCGCTCACTATTCACAGGTGAGAGACATCCACTATCTCTCAA 60
DB 951 ATGAAAGCGCGCTCATTACAGGTGAGAGACATCCACTATCTCTCAA 1010
QY 61 GCCACGAGCAG 120
DB 1011 GCCACGAGCAG 1070
QY 121 GCCATCAAGTTTCAACCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB 1071 GCTGTCAAGTTTCAACCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1130
QY 181 ATGATTAGCAATATTCAGTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB 1131 ATGATTAGCAATATTCAGTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1190
QY 241 GAAGTCAGAGCACTGACAGATACCCCTTACCAACAGAGAGAGAGAGAGAGAG 300
DB 1191 GATGTTAGGACCACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1250
QY 301 AATTAATACCTGGCAACATCCCTTTGATTTCTGTTCTCAGATGCAACAAGATTG 360
DB 1251 AATTAATACCTGGCAACATCCCTTTGATTTCTGTTCTCAGATGCAACAAGATTG 1310
QY 361 ATGAGTCGAGTTAGACACTGAGTGAAGTCACTATTCAGTATTTGGAAGAAATGAG 420
DB 1311 ATGAGTCGAGTTAGACACTGAGTGAAGTCACTATTCAGTATTTGGAAGAAATGAG 1370
QY 421 CTAGTAGACTCCCAAGATGCTGATCAATCAAGATTTTCACTACAGAGTCCACCG 480
DB 1371 CTAGTAGACTCCCAAGATGCTGATCAATCAAGATTTTCACTACAGAGTCCACCG 1430
QY 481 GGCCCGAGGGGTCCAGAGGTGACAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 1431 GGCCCGAGGGGTCCAGAGGTGACAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAG 1490
QY 541 AAGGACAGAAAG 600
DB 1491 AAGGACAGAAAG 1550
QY 601 CCAATTGAGACAGAGTGTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 1551 ACAATTGAGACAGAGTGTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1610

QY 661 GGCCCGAAAGGCTCCGTTGTTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 1611 GGTCCTCAAGAGATCTGTGTGTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1670
QY 721 CCAGGCCCCCGGGCCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 1671 CCAGGAGCCACAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1730
QY 781 TTCCAGAGAGTTCAG 840
DB 1731 TTCCAGAGAGTTCAG 1790
QY 841 GGCTTGCTGGGTGTACAG 900
DB 1791 GGCTTGCTGGGTGTACAG 1850
QY 901 TCAGAGAGAGT 960
DB 1851 TCAGAGAGAGT 1910
QY 961 GGCTTGCTGGGTGTACAG 1020
DB 1911 GGATGTCCAGCTCAG 1970
QY 1021 GAAATTTTGGAGATGCAAG 1080
DB 1971 GAAATTTTGGAGATGCAAG 2030
QY 1081 AACACTAG 1140
DB 2031 AACCTCAAG 2090
QY 1141 ATGCGCTCAG 1200
DB 2091 ATGCGCTCAG 2150
QY 1201 GACTACAAAAATTTGAAAG 1260
DB 2151 GATTACAAAAATTTGAAAG 2210
QY 1261 GAAGACTGTGCTGT 1320
DB 2211 GAAGACTGTGCTGT 2270
QY 1321 AATACTCTCACTTTGCGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371
DB 2271 AATACTCTCACTTTGCGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2321

RESULT 2
B0955927 957 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8763247 NIH_MGC_129 Mus musculus cdna IMAGE:6313500
DEFINITION 5', mRNA sequence.

ACCESSION B0955927
VERSION B0955927.1 GI:22371405
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>
AUTHORS 1 (bases 1 to 957)
TITLE Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
CDNA Library Preparation: Resgen, Invitrogen Corp.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1M13740 row: 1 column: 13
 High quality sequence stop: 665.
 Location/Qualifiers

FEATURES

source

1..957

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:6313500"
 /clone_lib="NIH_MGC_129"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Olfactory epithelium; Vector:
 pCMV-SPORT6.1.ccdp; Site_1: EcoRV; Site_2: NotI; Cloned
 unidirectionally. Primer: Oligo dT. Average insert size
 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this
 is a NIH_MGC library."

BASE COUNT 260 a 269 c 258 g 169 t 1 others
 ORIGIN

Query Match

Best Local Similarity 49.4%; Score 677.8; DB 14; Length 957;
 Matches 780; Conservative 0; Mismatches 113; Indels 5; Gaps 3;

QY 6 CAGCAGCTCAACTCATTTCACAGTCAGATGGAACATCACCACTATCTTCAGACCAA 65
 13 CCGGAGATCTACGCTCATTCACAGTCAGATGGAACATCACCACTATCTTCAGACCAA 72
 QY 66 CAGCAGAACCTGAAGAGCTGAGAGCTTACCAAGATGCGAGAAATGAACAGCCAT 125
 73 CCGAGAGACCTGAAGAGCTTACCAAGATGCGAGAAATGAACAGCCAT 132
 QY 126 CAGTTCACCAACTGAGAGAGCTTTCAGCTTTGAGAGCGATTTGTGAACATCAT 185
 133 CAGTTCACCAACTGAGAGAGCTTTCAGCTTTGAGAGCGATTTGTGAACATCAT 192
 QY 186 TAGCAATATCATTTACAGAGCCACCACTGCGAGCTGACCAAGCATCTAATGAAT 245
 193 TAGCAATATCATTTACAGAGCCACCACTGAGAGCATCTGACCAAGCATCTAATGAAT 252
 QY 246 CAGGAGCACTTGCACAGATACCTTACCAAGACACAGATGATCTGACCTCTTGAATTA 305
 253 TAGGAGCACTTGCACAGATACCTTACCAAGACACAGATGATCTGACCTCTTGAATTA 312
 QY 306 TACCTTGGCCCAACATCCGTTTGGATTCTGTTCTCAGAGATGACCAAGATTTGATGAG 365
 313 CACACTAGTCACATCCGTTTGGATTCTGTTCTCAGAGATGACCAAGATTTGATGAG 372
 QY 366 GTGAGGTTAGACACTGAAGTAGCCAACTTATCATGATTTGAGAAATGAAGTAGT 425
 373 GTCAAGTTAGACACTGAAGTAGCCAACTTATCATGATTTGAGAAATGAAGTAGT 432
 QY 426 AGACTCAGACATGTCACGCTCATCAAGATTTTACATATCTACAGGTCACCGGGCCC 485
 433 TGACTCAGACATGTCACGCTCATCAAGATTTTACATATCTACAGGTCACCGGGCCC 492
 QY 486 CAGGGGTCCAAAGAGTGACAGAGATCCAGAGGACCCCTGGCCCACTGCGCAACAAGG 545
 493 CAGAGGTCCAAAGAGTGACAGAGATCCAGAGGACCCCTGGCCCACTGCGCAACAAGG 552
 QY 546 ACAGAAAGAGAGAGGGGAGGCTTGACCACTGGCCCTGGGGTGAGAGAGCCCAAT 605
 553 ACAGAAAGAGAGAGGGGAGGCTTGACCACTGGCCCTGGGGTGAGAGAGCCCAAT 612
 QY 606 TGGACAGAGTGGTCCCGGAGAGAGGCTGGCGGCAAGATCTAAGGCTCCCAAGGGCCC 665
 613 TGGACAGAGTGGTCCCGGAGAGAGGCTGGCGGCAAGATCTAAGGCTCCCAAGGGCCC 672
 QY 666 CAAAGGCTCCGTTGTTCCCTGGGAGAGCCGCGCTTCAAGGAGCCCAAGTGGGAGCCAG 725
 673 CAAAGGATCTCGTGGGTCCCAAGGAGAGCTGGCCCTCAAGAGACCTAAGTGGGAGCCAG 732
 QY 726 CCGCCCGGGCCCAACAGGCAAGAGGAGTCCCGGCGCTCAGAGGCTCTGCTTCA 785

DB 733 ACCACAGGTCACACAGGAGAGATGAGTCCCTGGCCCTCANGGCCCTCTGCTT-CA 791
 QY 786 GGGACTTCAGAGGAGACCGTTGCGGAGCGCTGGAGCCCTCGGAGATG-CCAGGCT 844
 DB 792 GGGACTTCAGAGGAGACCGTTGCGGAGCGCTGGAGCGCTGGAGATG-CCAGGCT 851
 QY 845 TGGCTGGGGTACAGGACATGCTC--AGGCCCAAGGGGCCCCCGGCGCTCTGGAGCC 899
 DB 852 TGGCAGGGGTCCAGGACATGCTCCTGGGGCTTAAGGAGCACCTGGGCTCCAAAGGCC 909

RESULT 3
 BO891432 936 bp mRNA linear EST 16-AUG-2002
 LOCUS BO891432
 DEFINITION AGENCOURT_8754195 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6332754
 5', mRNA sequence.

ACCESSION BO891432
 VERSION BO891432.1 GI:22283446
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 TITLE NIH-MGC <http://img.ncl.nih.gov/>.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Mark Macconchie, Ph.D. and Nancy L. Freeman, Ph.D.

CDNA Library Preparation: Resgen, Invitrogen Corp
 cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1M13790 row: n column: 19
 High quality sequence start: 25
 High quality sequence stop: 632.
 Location/Qualifiers

FEATURES

source

1..936

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:6332754"
 /clone_lib="NIH_MGC_130"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: olocysts; Vector: pCMV-SPORT6.1.ccdp;
 Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
 Primer: Oligo dT. Average insert size 1.95 kb.
 Constructed by Resgen, Invitrogen Corp. Note: this is a
 NIH_MGC library."

BASE COUNT 273 a 250 c 244 g 163 t 6 others
 ORIGIN

Query Match

Best Local Similarity 42.4%; Score 581.2; DB 14; Length 936;
 Matches 667; Conservative 0; Mismatches 106; Indels 4; Gaps 2;

QY 1 ATGAACAGCCAGCTCAACTCATTTCACAGTCAGATGGAACATCACCACTATCTCTCAA 60
 131 ATGAATAGCAGCTCAGCTCATTCACAGTCAGATGGAACATCACCACTATCTCTCAA 190
 DB 61 GCCAAGAGCAGAGACTGAAAGCTCGAGAGATTTACCAAAAGATGACAGATTAACA 120
 191 GCCAAGAGCAGAGCTGAAAGCTTCAGAGATTTACCAAAAGATGACAGATTAACA 250
 QY 121 GCCATCAAGTTCAACCACTGAGAGAGCTTTCACACTCTTTGAGAGGATATTGTGAAC 180
 251 GCGTCAAGTTCAACCACTGAGAGAGCTTTCACACTCTTTGAGAGGATATTGTGAAC 310
 QY 181 ATCATTAGCAATATTCATACAGGCAACACCTCGGAGCGTGAACAGCAATCTTAAT 240

Db 311 ATCATTTGCAACATCACTACACAGCCCATCAGTACAGACATCTGTAAT 370
QY 241 GAAGTCAGACCACTTCCAGATACCTTACCAAAACACAGATGATCTGCTCTG 300
Db 371 GATTTAGGACACATGACACAGACCTTGACACAGACAGATGACCTGCTCTG 430
QY 301 AATTAATACCTGGCCAAACATCTGTTGATTTCTCTCAGATGACAAAGATTG 360
Db 431 AATTAACACTAGTCAACATCTGCTGATTTCTCTCAGATGACAAAGATG 490
QY 361 ATGAGTCGAGGTAGACACTAGTACCACTATCAGTATGATGAAAGATGAG 420
Db 491 ATGAGTCGAAAGTATGACACTGAGTGGCCACTTATCAGTATGATGAAAGATG 550
QY 421 CTAGTACCTCCAGACATGCTGACCTCATCAAGATTTTACATATCTACAGGTCACCG 480
Db 551 CTGCTGACCTCCAGACAGGTCAGCTCATCAAGATTTTACATATCTACAGGTCACCG 610
QY 481 GGGCCCAAGGGGTCACAGATGACAGAGATCCAGGAGCCCTGGCCCACTGGCAAC 540
Db 611 GGGCCCAAGGGGTCACAGAGATCTCAGAGGACCTGGCTCCCACTGGCAAC 670
QY 541 AAGGACAGAAAGAGAGAGGGGGGCTGACACCTGGCCCTGGGGGAGAGAGGC 600
Db 671 AAGGACAGAAAGAGAGAGGGAGAGCTGCTCCTGCTGCTGGGGAGAGAGGC 730
QY 601 CCAATTGACACAGCTGCTCCCTCCGAGAGCTGGCGCAAGAGATCTAAAGCTCCAG 660
Db 731 ACATTTGACACAGTCCGCTCTGAGAGAGGTCGACAAAGATCTCAAGGCTCAGAG 790
QY 661 GGGCCCAAGG--CTGCCGTGTTCCCTTGGGAAGCCGGCCCTCAGGCCCCAGTGGG 718
Db 791 GGNCTCAAGAGATCTGCTGGTCCCTCCAGGAGAGCTGGCCCTCAAGGGGCTGCTGGG 850
QY 719 ACCCAGG--CCCCCGGCGCACCAAGGAGACTCCCGGGCCCTCAGGCCCC 773
Db 851 ACCCAGGACCACTGCTCCACACAGGAGGATGACTCCCTGGGCTCAGGAGGC 907

RESULT 4
BI456109 906 bp mRNA linear EST 21-AUG-2001
LOCUS 603172765F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:525188 5',
DEFINITION mRNA sequence.
ACCESSION BI456109
VERSION BI456109.1 GI:15246765
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 906)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@emall.nih.gov
Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
[found through the I.M.A.G.E. Consortium/LLNL at:](http://image.llnl.gov)
<http://image.llnl.gov>
Plate: LHAM1636 row: b column: 17
High quality sequence stop: 730.
Location/Qualifiers
1..906
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:525188"
/clone_lib="NCI_CGAP_Mam5"

FEATURES
source
1..906
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:525188"
/clone_lib="NCI_CGAP_Mam5"

/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SpOrf6; Site: 1; Salt:
site: 2; Not: Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lohar Hennighausen/Robin Humphreys,
NIH"
BASE COUNT 247 a 228 c 256 g 175 t
ORIGIN

Query Match 40.8%; Score 559; DB 13; Length 906;
Best Local Similarity 82.4%; Pred. No. 3.7e-127;
Matches 703; Conservative 0; Mismatches 140; Indels 10; Gaps 5;
QY 224 TGACACCAATCTGAATGAAGTACGACCACTTGACAGATACCTTACCAACACAG 283
Db 1 TGACACCAATCTGAATGAAGTATGATGATGATGATGATGATGATGATGATGATG 60
QY 284 ATGATCTGACCTCTGATTAATTAATACCTTGCCCAACATCCCTTGATCTGTTCTCA 343
Db 61 ATGACCTGACCTCTGATTAATTAATACCTTGATTAATTAATTAATTAATTAATTA 120
QY 344 GATGACCAAGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATG 403
Db 121 GATGACCAAGATTTTATGATGATGATGATGATGATGATGATGATGATGATGATG 180
QY 404 TTATGGAAGAATGAACCTGTGATGATGATGATGATGATGATGATGATGATGATG 463
Db 181 TTATGGAAGAATGAACCTGTGATGATGATGATGATGATGATGATGATGATGATG 240
QY 464 TACTACAAGTTCACCGGGGCCAGGGGTCACAGAGTGAAGATGCCAGGAGCC 523
Db 241 TTCTACAAGTTCCTCTGCGCCCGAGAGTTCACAGAGTGAAGATGCCAGGAGCC 300
QY 524 CTGGCCCACTGCGACAAAGAGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 583
Db 301 CTGGTCCCACTGCGACAAAGAGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 584 CTGGGGTGAAG 643
Db 361 CTGGGGTGAAG 420
QY 644 GATCTAAAGGCTCCAGAGGCCCCCAAGGCTCCGTTGTTCCCTGGAAGCCCGCCCTC 703
Db 421 GATCTAAAGGCTCCAGAGGCCCCCAAGGATCTCGTGGGTCCCAAGTGAAGCTCGGCC 480
QY 704 AGGCCCCAGTGGGAGACCCAGGCCCCCGGGCCACAGCAAGAGAGACTCCCGGCC 763
Db 481 AAGGACCTAGTGGGAGACCCAGGCCCCAGGCTCCAGGAGAGAGAGAGAGAGAGAG 540
QY 764 CTCAGGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 823
Db 541 CTCAGGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
QY 824 GACCTGGGGAGCTGCCAGGCTTGCTGGGTGCTTACAGGATGCGAGGCCCAAGGGCCCC 883
Db 601 GACCTGGGGAGCTGCCAGGCTTGCTGGGTGCTTACAGGATGCGAGGCCCAAGGGCCCC 660
QY 884 CCGGCTCTCTGCGCCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 943
Db 661 TAGG-CCTTCAGGAGGCTTCCAGGAGC-ATGGAAGCTATGCTGTCAGAAAGAGAGAG 718
QY 944 CGGACCGGAGAGCAATGAC--TGCCCGCTCAGCTGGAAGCAATTCACA---GACAAATG 998
Db 719 CAGCATAGAGGTTCAACGCGATGTCGCTCAGTGAAGAACTTTCAGGATTAATG 778
QY 999 CTACTATTTTCAAGTGG--AGAAAGAAATTTTGAAGATGCAAGCTTTCTGTAAGA 1055
Db 779 CTACTATTTTCAATGCAAAAGAAATTTTGAAGATTTGTAAGCGTTCTGTGAGA 838
QY 1056 CAAGTCTTCAT 1068
|| | ||| ||

| Db | 839 | AAATTTTCCAAT | 851 |
|------------|---|--------------|---------------------------------------|
| RESULT_5 | | | |
| LOCUS | BO713873 | | |
| DEFINITION | BO713873 | 861 bp | mRNA |
| ACCESSION | AGNCNCURT_8482828 | NIH_MGC_129 | Mus musculus CDNA clone IMAGE:6306240 |
| VERSION | 5, | | mRNA sequence. |
| KEYWORDS | BO713873 | | |
| SOURCE | BO713873.1 | GI:21852772 | EST. |
| ORGANISM | house mouse. | | |
| | Mus musculus | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| REFERENCE | 1 (bases 1 to 861) | | |
| AUTHORS | NIH-MGC http://mgc.ncl.nih.gov/ . | | |
| TITLE | National Institutes of Health, Mammalian Gene Collection (MGC) | | |
| JOURNAL | Unpublished (1999) | | |
| COMMENT | Contact: Robert Strausberg, Ph.D. | | |

```

Plate: L1AM13721 row: n column: 01
High quality sequence stop: 598.
FEATURES
    Location/Qualifiers
        1.. 861

```

| BASE COUNT | ORIGIN |
|----------------------------------|--|
| 221 a 253 c 239 g 147 t 1 others | /organism="Mus musculus" /db.xref="taxon:10090" /clone="IMAGE:6306240" /clone_id="NIH_MGC_129" /lab_host="DH10B (phage-resistant)" /note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1.ccd; Site.1: EcoRV; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt; Average insert size 2.2 kb. Constructed by Reseen, Invitrogen Corp. Note: This is a NIH_MGC Library." |

| | | | | | | |
|-------------|--------|-------|--------|--------|--------|------|
| Query Match | 40.6%; | Score | 556.4; | DB 14; | Length | 861; |
|-------------|--------|-------|--------|--------|--------|------|

Best Local Similarity 82.8%; Pred.NO. 1.0E-126;
Matches 647; Conservative 0; Mismatches 132; Indels 2; Gaps 1.

359 TGATGAGTCGAGGTTAGACACTGAAGTAGCCAACTATCAGTCATTATGGAAGAATCA 418

db 14 TGATGAGGTCAAAGTTAGACACTGAAGTGGCCACTTATCAGTGGTTATGGAAGATGCA 73

419 AGCTAGTACACTCCAAGCATGGTCAGCTCATCAAGAAATTTACATACTACAAAGGTCCAC 478

Db 74 AACGTGTTGACTCCAGACGCGGTACGTCATCAGAACTTTACCATTTACAGAGTCCTC 133

479 CCGGGCCCAAGAGGTGACAGAGGATCCCAAGGACCCCTGAGCCCAACTGGCA 538

Db 134 C T G G C C C A G A G T C C A A A G G T G A C A G A G G A T C T C A G G G A C C T G T G T C C A A C T G G C A 193

0v 539 ACAAGGGACAAAGSAGSAGGAGGCCCTGACCCTGGCCCTCGGSGAAGAN 598

Dh 194 ACAAGGACAGAAAGACGAAGCCCTGCATTCACCTCGGCCCTGGCAGAACG 253

[illegible][illegible]

DD 234 GCACAAATGGACCGAGTCGGCCCTCTCGAGAGAGCGTGCGACGACAAAGGATCCAAAGGCTCAC 313

659 AGGGCCCCAAGGCTCCCGTGTTCCCTGGAGACCGGGCCCTCAGGGCCCAAGTGGGS 718

Db 314 AGGGTCCCAAGGATCTGTGGGTCCCCAGGGAAGCCTGGCCCTCAGGACTAGTGGGG 373

[illegible]

| | |
|------------|--|
| RESULT | 6 |
| LOCUS | B0674807 |
| DEFINITION | B0674807 668 bp mRNA linear EST 15-JUL-2002 |
| ACCESSION | AGNCOURT_8035172 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6212672 |
| VERSION | B0674807 |
| KEYWORDS | B0674807.1 GI:21785641 |
| SOURCE | EST. |
| ORGANISM | human. |
| | Homo sapiens |

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Mammalia; Chordata; Crinacea; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 668)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

```

Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MCC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LCM2377 row: C column: 09
High quality sequence stop: 656.
Location/Qualifiers
1..668
FEATURES
source
```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6212672"
/clone_id="NH_MGC_102"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAC(G) library constructed by Ling Hong in the Laboratory of Gerald M. Rubin

```

(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH-MGC Library."

BASE COUNT 216 a 148 c 167 g 137 t
ORIGIN
Query Match 38.3%; Score 525.4; DB 14; Length 668;
Best Local Similarity 99.8%; Pred. No. 6,4e-119;
Matches 526; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 845 TGGCTGGGATCCAGGATGCGAGCCGCCCAAGGCCCCCGGCTCTCTGCCCCATCAG 904
|||||
Db 1 TGGCTGGGATCCAGGATGCGAGCCGCCCAAGGCCCCCGGCTCTCTGCCCCATCAG 60
QY 905 GAGCGGTGGTCCCTGGCCCTGCGAATGAGCCAGCCCGGCGCCGAGAGACATGCT 964
|||||
Db 61 GAGCGGTGGTCCCTGGCCCTGCGAATGAGCCAGCCCGGCGCCGAGAGACATGCT 120
QY 965 GCGCCGCTCAGTGAAGAACTTCACAGACAATGCTACTATTTTTCAGTTGAGAAAGAA 1024
|||||
Db 121 GCGCGCTCAGTGAAGAACTTCACAGACAATGCTACTATTTTTCAGTTGAGAAAGAA 180
QY 1025 TTTTGGAGTGCAGAACTTTCTGTGAAGACAGCTTCACATCTTTTCTAATAACA 1084
|||||
Db 181 TTTTGGAGTGCAGAACTTTCTGTGAAGACAGCTTCACATCTTTTCTAATAACA 240
QY 1085 CTAGAGAGGAACAGCAATGATTAATAAAGAGATGTAGGAGAGAGAGCCACTGATCG 1144
|||||
Db 241 CTAGAGAGGAACAGCAATGATTAATAAAGAGATGTAGGAGAGAGAGCCACTGATCG 300
QY 1145 GCGTCAACAGCTCAGACCGTGAATAATGAATGAAAGTGGCTGATGAGACATCTCCAGACT 1204
|||||
Db 301 GCGTCAACAGCTCAGACCGTGAATAATGAATGAAAGTGGCTGATGAGACATCTCCAGACT 360
QY 1205 ACAGAAATTTGAAAGCTGAGACAGCCGATTAAGTGGGTCATGGCCATGGGCGAGAGAG 1264
|||||
Db 361 ACAGAAATTTGAAAGCTGAGACAGCCGATTAAGTGGGTCATGGCCATGGGCGAGAGAG 420
QY 1265 ACTGTGCTGGGTTGATTTATGCTGGGCACTGGAAGCATTTCCAAATGTGAAGACCTCATTA 1324
|||||
Db 421 ACTGTGCTGGGTTGATTTATGCTGGGCACTGGAAGCATTTCCAAATGTGAAGACCTCATTA 480
QY 1325 ACTTCATTTGCGAAAAAGACAGGAGACAGTCTGTCATCTGCATTA 1371
|||||
Db 481 ACTTCATTTGCGAAAAAGACAGGAGACAGTCTGTCATCTGCATTA 527

RESULT 7
AL568743/c 967 bp mRNA linear EST 16-FEB-2001
LOCUS AL568743 LTI_FL002_PL1 Homo sapiens cDNA clone CS0DE005YH04 3 prime
DEFINITION ' mRNA sequence.
ACCESSION AL568743
VERSION AL568743.1 GI:12923387
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 967)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polyes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..967
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DE005YH04"
/clone_lib="LTI_FL002_PL1"

/lab_host="DH10B"
/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dt) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 222 a 221 c 217 g 290 t 17 others
ORIGIN
Query Match 34.7%; Score 475.4; DB 9; Length 967;
Best Local Similarity 96.7%; Pred. No. 1.7e-106;
Matches 504; Conservative 3; Mismatches 9; Indels 5; Gaps 2;

QY 851 GGGTACCAAGGCGATGCCAGGCCCCCAAGGGCCCCCGGCTCTCTGCCCCATCAGAGCGG 910
|||||
Db 898 GGGTACCAAGGCGATGCCAGGCCCCCAAGGGCCCCCGGCTCTCTCTGCCCCATCAGAGCGG 843
QY 911 TGGTGGCCCTGGCCCTGCGAATGAGCCAGCCCGGCGCCGAGAGACATGCTGCCCGC 970
|||||
Db 842 TGGTGGCCCTGGCCCTGCGAATGAGCCAGCCCGGCGCCGAGAGACATGCTGCCCGC 783
QY 971 CTCACGTGAAGAACTTCACAGACAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTG 1030
|||||
Db 782 CTCACGTGAAGAACTTCACAGACAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTG 723
QY 1031 AGGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTTTTCATATAACACTAGAG 1090
|||||
Db 722 AGGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTTTTCATATAACACTAGAG 663
QY 1091 AGGAAACAGCAATGATTAATAAAGAGATGTAGGAGAGAGAGCCACTGATGGCCCTCA 1150
|||||
Db 662 AGGAAACAGCAATGATTAATAAAGAGATGTAGGAGAGAGAGCCACTGATGGCCCTCA 603
QY 1151 CAGACTCAGAGCCGGAATAATGAATGAAGTGGCTGGATGGACATCTCCAGACTACAAAA 1210
|||||
Db 602 CAGACTCAGAGCCGGAATAATGAATGAAGTGGCTGGATGGACATCTCCAGACTACAAAA 543
QY 1211 ATTGGAAGCTGAGACAGCCGATTAAGTGGGTCATGGGCCATGGAGAAAGACTGTG 1270
|||||
Db 542 ATTGGAAGCTGAGACAGCCGATTAAGTGGGTCATGGGCCATGGAGAAAGACTGTG 484
QY 1271 CTGGGTTGATTTATGCTGGGCGAGTGAAGATTTCCAAATGTGAAGAGCTCAATTAACCTCA 1330
|||||
Db 483 CTGGGTTGATTTATGCTGGGCGAGTGAAGATTTCCAAATGTGAAGAGCTCAATTAACCTCA 424
QY 1331 TTTCGGAAGAAAGACAGGAGACAGTACTGTCAATCTGCATTA 1371
|||||
Db 423 TTTCGGAAGAAAGACAGGAGACAGTACTGTCAATCTGCATTA 383

RESULT 8
BE910803 580 bp mRNA linear EST 29-SEP-2000
LOCUS BE910803 60161855F1 NCL_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3962292 5',
DEFINITION 'mRNA sequence.
ACCESSION BE910803
VERSION BE910803.1 GI:10407765
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 580)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1AM9128 row: m column: 13
 High quality sequence stop: 580.

FEATURES

source

Location/Qualifiers

1..580

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:3962292"

/clone_lib="NCI-GCAP_Mam1"

/tissue_type="tumor, biopsy sample"

/dev_stage="10 months, virgin"

/lab_host="DH10B"

/note="Organ: mammary; Vector: PCMV-SPORT6; Site:1; Salt:

Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigator

providing samples: Gilbert Smith, NIH"

BASE COUNT 159 a 153 c 161 g 107 t

ORIGIN

Query Match

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

Best Local Similarity 32.5%; Score 445.6; DB 12; Length 580;

Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Query Match 32.5%; Score 445.6; DB 12; Length 580;

LOCUS AW958053 552 bp mRNA linear EST 01-JUN-2000
 DEFINITION EST370123 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
 ACCESSION AW958053
 VERSION AW958053.1 GI:8147736
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 552)
 Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspar,R., Gay,C., Holt
 ,I.E., Saeed,A.I., Sharov,Y., Lee,N.H., Yeatman,T.J. and
 Quackenbush,J.
 Assessment of gene expression patterns in a model of colon tumor
 metastasis using a 19,200 element cDNA microarray
 Unpublished (2000)
 CONTACT: John Quackenbush
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 3528
 Fax: 301 838 0208
 Email: john@igf.org
 Plate: 115
 Seq primer: Reverse.
 Location/Qualifiers

FEATURES

source

1..552

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="MAGE resequences, MAGE"

/note="Vector: pBluescriptSkm"

BASE COUNT 176 a 112 c 143 g 121 t

ORIGIN

Query Match

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Query Match 31.9%; Score 437.2; DB 10; Length 552;

Best Local Similarity 31.9%; Score 437.2; DB 10; Length 552;

| | | | | | | |
|---------------------------|--|--|-----|--|--|--|
| DEFINITION | 6010892246t1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3484163 5' | | | | | |
| ACCESSION | BE290289 | | | | | |
| VERSION | BE290289.1 GI:9171250 | | | | | |
| KEYWORDS | EST. | | | | | |
| SOURCE | house mouse. | | | | | |
| ORGANISM | Mus musculus | | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 601) | | | | | |
| AUTHORS | NIH-MGC http://mgc.nci.nih.gov/. | | | | | |
| TITLE | National Institutes of Health, Mammalian Gene Collection (MGC) | | | | | |
| JOURNAL | Unpublished (1999) | | | | | |
| COMMENT | Contact: Robert Strusberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINT) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. consortium/LINT at: http://image.llnl.gov Plate: LLM8516 row: k column: 12 High quality sequence stop: 551. Location/Qualifiers 1..601 | | | | | |
| FEATURES | source organism="Mus musculus" strain="C57BL/6J" db_xref="taxon:10090" clone="IMAGE:3484163" clone_lib="NCI CGAP Mam5" tissue_type="tumor, gross tissue" dev_stage="7 months" lab_host="DH10B" note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH" | | | | | |
| BASE COUNT | 156 a 161 c 179 g 105 t | | | | | |
| ORIGIN | | | | | | |
| Query Match | 30.3%; Score 415.8; DB 10; Length 601; | | | | | |
| Best Local Similarity | 84.8%; Pred. No. 7.3e-92; | | | | | |
| Matches 501; Conservative | 0; Mismatches 87; Indels 3; Gaps 3 | | | | | |
| Dn | 541 | AAGGACAGAAGAGAGAGAGGGGAGCGCTTGACCACTGGCCCTCGGGTGAGAGAGGC | 600 | | | |
| | 1 | | 111 | | | |
| | 9 | AGGTGACAAAGAGAGAGAGAGAGAGCGCTGTCACACTCGGCCCTCGGGTAGAGAGGGC | 68 | | | |
| Dn | 601 | CCAATTGGACCACTGCTGCTCCCCCGGAGAGCGTGGCGGCAAGAGATCTAAGGCTCCAC | 660 | | | |
| | 69 | ACAATTGGACCACTGCGCCCTCTCTGGAGAGCGTGGCAGCAAGAAAGATCCAAAGGCTCAC | 128 | | | |
| Dn | 661 | GGCCCCAAGAGGCTCCCGTGGTTCCCTCGG -AAGCCCGGCGCTCAGGGGCCCAAGTGGGA | 719 | | | |
| | 129 | G | 188 | | | |
| Dn | 720 | CCCAGCCCCCGGGGCCACAGGCAAGAGGAGCTCCCGGCGCTCAGGGGCCCTCTCGG | 779 | | | |
| | 189 | | 248 | | | |
| Dn | 780 | CCTTCCAGGACCTTCAGGGCACCCTTGGGAGCCTTGGGAGCTCTGACCTCGGGGACTGCC | 839 | | | |
| | 249 | | 308 | | | |
| Dn | 840 | AGGCTTGCTGGGGTACAGGCACTGCCAGGCCCAAGGGCCCCCGGCGCTCTCGGCC | 899 | | | |
| | 309 | AGGCTTGCT -AAGGCTCCAGGCACTGCTCTGGGCTTAAGGGGACCACTGCGCCCTCAGGCC | 367 | | | |
| Dn | 900 | ATCAGAGAGGAGTGTGCTCCCTCGGCTCTGAGATGAGCAACCCCGGCAACCGGAGAGCAA | 959 | | | |
| | 368 | CTTCAGAGCAATTCAGGACCTTGGCTCTTGAGATGAGCAACCCCAAGCATCAGAGTGA | 427 | | | |

| QY | 960 | TGGCGGCCGCTCAGCTCGGAGAACTTCACAGACAAATGCTATTTTTCAGTTGAGAA | 1019 |
|-----------------------|--|--|----------------------------------|
| Db | 428 | CGGATGTCGGCTCCTCTGGAAGAACTTCACGATATAATGCTACTATTTTCTTGGAAAA | 487 |
| QY | 1020 | AGAAATTTTGTAGAGTGCAGAAAGCTTTTGTGTGGAAGACAGCTCTTCACATCTGTTTCAT | 1079 |
| Db | 488 | AGAAA-TTTTGAAGATGCTGAAGCTTTCTGTGTGAAGACAAATCTTCCATCTGTTTCAAT | 546 |
| QY | 1080 | AAACACTGAGAGACAGCAGCAATGATATAAAAAACAGATGCTAGGAGAGAA | 1130 |
| Db | 547 | AAACTCAAGAGAGAAAGAACAGCAATGTGGTATAAAAAACGATACGTGGGGAGAGA | 597 |
| RESULT 11 | | | |
| LOCUS | B0934501 | 884 bp | mRNA linear EST 21-AUG-2002 |
| DEFINITION | AGENCOURT.8765665 NIH_MGC_130 | Mus musculus | cdna clone IMAGE:6330794 |
| ACCESSION | B0934501 | | |
| VERSION | B0934501.1 | GI:22349884 | |
| KEYWORDS | EST. | | |
| SOURCE | house mouse. | | |
| ORGANISM | Mus musculus | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| AUTHORS | 1 (bases 1 to 884) | | |
| TITLE | NIH-MGC http://mgc.nci.nih.gov/. | | |
| JOURNAL | National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) | | |
| COMMENT | Contact: Robert Strausberg, Ph.D. Email: cga@bbs-rt@mail.nih.gov Tissue Procurement: Mark Macconochie, Ph.D. and Nancy L. Freeman, Ph.D. | | |
| FEATURES | CDNA Library Preparation: Resgen, Invitrogen Corp CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: L1AM13785 row: m column: 03 High quality sequence stop: 352. Location/Qualifiers 1..884 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:6330794" /clone_lib="NIH_MGC_130" /lab_host="DH10B (phage-resistant)" /note="Organ: otcysts; Vector: pcMV-SPORT6.1.ccd; Site:1: EcoRV; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.95 kb. Constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC Library" | | |
| BASE COUNT | 282 a | 225 c | 205 g |
| ORIGIN | | | 172 t |
| Query Match | 28.1%; | Score 385.6; | DB 14; Length 884; |
| Best Local Similarity | 82.9%; | Pred. No. 2.5e-84; | |
| Matches | 489; | Conservative 0; | Mismatches 94; Indels 7; Gaps 4; |
| QY | 1 | ATGAACAGCCAGCTCAACTCATTCATTCACA-GGTCAGATGAGAAATCACCACATATCTCTCA | 59 |
| Db | 277 | ATGATATACCAAGCTCAGCTCATTCACAGGGTCAGATGAGCAACATATACCATATCTTCACA | 336 |
| QY | 60 | AGCCACAGACGACGAACCTGGAAGACCTCGACAGACTTACACAAAGATGCAGAAATATGAAAC | 119 |
| Db | 337 | GGCCAAGACGACGAGACCTGGAAGACCTTCAGGACTTACCAAGGATATACGAAATATGAAAC | 396 |
| QY | 120 | AGCCATCAAGTTTCAACCAACTGTGAGAGAACGCTTCACAGCTCTTGTGAGACGATATTTGTGAA | 179 |
| Db | 397 | AGCTGTCAAGTTTCAAGCCCAACTTTGAGGAACGCTTCCAGGCTTTGTGAGACAGATATTTGTGAA | 456 |

| | | | |
|------------|--------------|---|-----------------------------|
| Y | 180 | CATTTTGGCAATATACAGTTTACACAGCCACACCTGGGAGCGCTGACCAATCTAA | 239 |
| DB | 457 | CATCATTTGGCAATATACAGTTTACACAGCCACACCTGGGAGCGCTGACCAATCTAA | 516 |
| Y | 240 | TGAATCTGAGGACCATCTTGACAGATACCTCTTACCAAAACACAGATGATCTGACCTCT | 299 |
| DB | 517 | TGATTTTGGAGACCATGACATGACAGACACCTTGACACACACGATGATGACCTCTCT | 576 |
| Y | 300 | GAAATATACCTTGGGCAACATCTCCGTTTGGATTCTGTTTCTCTCAGATGCAACAGATT | 359 |
| DB | 577 | GAAAT-ACACACTATGACATCATCTCCGTTTGGATTCTATTTCTCTCAGATGCAACAGAT | 635 |
| Y | 360 | GATGAGGTCGAGGTTAGACACTGTAAGTACCCACTTTATCATGATTTATGGAAGAATGA | 419 |
| DB | 636 | GATGAGGTCGAGGTTAGACACTGTAAGTACCCACTTTATCATGATTTATGGAAGAATGA | 695 |
| Y | 420 | GCTATGATGACCTTCCAGACATGCTGACCTCATCAAGAAATTTTCAATGATCAAGGTCAC | 479 |
| DB | 656 | ACTGTTGATCTTCAAGCAGCGTGCACCTCATCAAGAAATTTTCAATGATCAAGGTCAC | 755 |
| Y | 480 | GGGCGCCGAGGGTCCCAAGAGTGTT--ACAGAGGATATCCAGGAGCCCTCTGG--CCCACT | 534 |
| DB | 756 | TGGCCCGAGAGGTCAAAAAGGGGAGACAGAGGATATCTCAGGGAGCAACTGGTCCAACTT | 815 |
| Y | 535 | GGCAACAAGGACAGAAAGAGAGAGAGGGGAGGAGCCTGTGACCACTGGCC | 584 |
| DB | 816 | GGCAACAAGGAGGAGACAAAAGGAGGAGAGAGAGGGGAGAGCCTGTGTCTC | 865 |
| RESULT 12 | | | |
| LOCUS | BB248064 | 638 bp | mRNA linear EST 23-Oct-2001 |
| DEFINITION | BB248064 | RIKEN full-length enriched, 7 days neonate cerebellum Mus | |
| ACCESSION | BB248064 | | |
| VERSION | BB248064.2 | GI:16355610 | |
| KEYWORDS | EST. | | |
| SOURCE | house mouse. | | |
| ORGANISM | Mus musculus | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| COMMENT | | | |

Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)

wag1, K., Fujiwara, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E.,
Watabiki, M., Yoneda, Y., Ishikawa, T., Okawa, K., Tanaka, T., Matsura
, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multipillar sequencer. Genome Res.
10 (11), 1757-1771 (2000).

| | |
|-----------------------|--|
| | Kono H., Fukunishi Y., Shibata K., Itoh M., Carninci P., Sugahara Y., and Hayashizaki Y. |
| | Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11(2), 281-289 (2001) |
| | Kondo S., Shingawa A., Saito T., Kiyosawa H., Yamanae I., Alzawa K., Fukuda S., Hara A., Itoh M., Kawai J., Shibata K. and Hayashizaki Y. |
| | Computational Analysis of Full-Length Mouse CDNA's Compared with Human Genome Sequences Mamm. genome. 12, 673-677 (2001) |
| | Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. |
| FEATURES | cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. |
| source | Location/Qualifiers 1..638 organism="Mus musculus" /db_xref="taxon:10090" /clone="A730023E20" /clone_lib="RIKEN full-length enriched, 7 days neonate cerebellum" /tissue_type="cerebellum" /dev_stage="7 days neonate" /lab_host="DH10B" /note="Site_1: SalI; Site_2: BamHI: cDNA Library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGACGAGAAGCATCCAAAGCGCTCTTTTATTTTTTTVN 3'] cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTTCGGACTTAATAATAATAATCAAAAACCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified Bluescript KS(+) after bulk excision from lambda da |
| BASE COUNT | 212 a 134 c 146 g 146 t |
| ORIGIN | |
| Query Match | 26.9%; Score 369.4; DB 10; Length 638; |
| Best Local Similarity | 84.5%; Pred. No. 2.2e-80; |
| Matches 415, | Conservative 0; Mismatches 76; Indels 0; Gaps 0; |
| OY | 881 CCCC GGCCGC CTGCC ATCAG AGAG CGGG TGCT GCC CCT GC CC GT CA GA TA TC GC AA 940 |
| Dd | 1 CAC CTG GCG CTC CA AG CCC CT TAG AGA CAT GC ACC ATT GGT CTG CA GTA TC AAC CA 60 |
| OY | 941 CCCC GC AC CGG AG AGACA TGG CTG CCC GC CT CAC TTGGA AGAAC TTC ACA CAG ACAA NTG CT 1000 |
| Dd | 61 CCC CAG CAA TC AAG SATG CCA AGSG ATTC CGG CCT CAC ATG GA AAG ACTT CAC AAG ATA ATG CT 120 |
| OY | 1001 ACT ATTTT TCA GTTG AG AAA CA AAT TTTT TGA GAG ATG CAAA GCT TTT TGT GGA GCA AGT 1060 |
| Dd | 121 ACT ATTTT TCA GTT GGA AAAA AGAAA ATTTT TGA AAG ATG CTA AGCT TTT TGT GGA GCA AAT 180 |
| OY | 1061 CTT CAC ATC TTT GTT TCA TPA AAC COT PAC AGAG GAA GACC ANT GST ATTA AAAAAA CAG AN TG 1120 |
| Dd | 181 CTT CAC ATC TGT TTT TCA TPA AAC CTA PA AGA AGAAC CACA AATG ATG ATTA AAAAAA GCA TTA CGG 240 |
| OY | 1121 TAG GAG AGAG AGA GGC CAC ATG ATG GGC CTC CACA GAT CAC AGC GTT GA AATG AATG AAGT 1180 |
| Dd | 241 TGG GAG AGA GAA AAG CCA TTG CAT GGG CCT CACA GAC TCA GAA CAG AAG AAG CCA AATG AAGT 300 |
| OY | 1181 GGC TGG ATG GAC ATC TCC AGA CTA CAAA AAT TGG AAA AGCTG GAC AGC CGA TAA CTGGG 1240 |

Db 301 GGCATGACGGGTCACCTGTTGATTAACAAACTGGAAGCTGACACCAACATCACTGGG 360
 QY 1241 GTCATGCGCCATGGCCAGGAGAAAGACTGTGCTGGCTGATTTATGCTGGCAGTGAAGC 1300
 Db 361 GGAGTGGCCATGGCCAGGAGAAAGACTGTGCTGGCTGATTTATGCTGGCAGTGAAGC 420
 QY 1301 ATTTCGAATGTGAAGAGCTCAATTAACCTGATTTGGCAAAAAGACGAGACACTGT 1360
 Db 421 ACTTCAGATGTGAATCAATTAACCTGATTTGGCAAAAAGAGGAGGAGCAGTACCAT 480
 QY 1361 CATCTGCATTA 1371
 Db 481 CATCCATTTTA 491

RESULT 13
 BM676508/c 500 bp mRNA linear EST 27-FEB-2002
 LOCUS BM676508
 DEFINITION UI-E-EJ0-ahq-c-05-0-UI.s2 UI-E-EJ0 Homo sapiens cDNA clone
 ACCESSION BM676508
 VERSION BM676508.1 GI:18986404
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 500)
 AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@iuii.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Forward

FEATURES

source

Location/Qualifiers
 1..500

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-EJ0-ahq-c-05-0-UI"

/clone_lib="UI-E-EJ0"

/tissue_type="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina foveal and Macular, RPE and
 Choroid"

/dev_stage="fetal and adult"

/lab_host="PH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-EJ0 is a subtracted cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT73-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for this library are: fetal eyes, AGAATCAGA
 ; lens, CGATTACGCA; eye anterior segment, AATGCCGCAT;
 optic nerve, CCATTACGTG; retina, CCGCG; Retina foveal and

Macular, GTCG; RPE and Choroid, ACCCA. This library was
 created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI).
 TAG_LIB=UI-E-EJ0
 TAG_TISSUE=human eye anterior segment
 TAG_SEQ=AAATGCCCAT"
 BASE COUNT 109 a 123 c 89 g 179 t
 ORIGIN

Query Match 25.6%; Score 351.4; DB 14; Length 500;
 Best Local Similarity 99.7%; Pred No. 5.4e-76;
 Matches 352; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1019 AAGAAATTTTGGAGATGCAAGCTTTTCTGTGAACAAGCTTTCACATCTGTTTCA 1078
 Db 494 AGGAATTTTGGAGATGCAAGCTTTTCTGTGAACAAGCTTTCACATCTGTTTCA 435
 QY 1079 TAAACACTAGAGAGGAGCAATGATTAATAAACAGTGTAGGAGAGAGCCACT 1138
 Db 434 TAAACACTAGAGAGGAGCAATGATTAATAAACAGTGTAGGAGAGAGCCACT 375
 QY 1139 GGATCGGCTCACAGACTGAGACCGTGAATAATGAATGAGTGGCTGATGGACATCTC 1198
 Db 374 GGATCGGCTCACAGACTGAGACCGTGAATAATGAATGAGTGGCTGATGGACATCTC 315
 QY 1199 CAGACTACAAATTTGGAAAGCTGACACCCGGATTAATGAGGCTCATGGGCCAG 1258
 Db 314 CAGACTACAAATTTGGAAAGCTGACACCCGGATTAATGAGGCTCATGGGCCAG 255
 QY 1259 GAGAACATCTGCTGGCTGATTTATGCTGGCAGTGAACGATTTCCAAATGTGAAGAG 1318
 Db 254 GAGAACATCTGCTGGCTGATTTATGCTGGCAGTGAACGATTTCCAAATGTGAAGAG 195
 QY 1319 TCATTAACCTTCAATTTGCGAAAAAGACAGGAGACACTGTCATCTGATTA 1371
 Db 194 TCATTAACCTTCAATTTGCGAAAAAGACAGGAGACACTGTCATCTGATTA 142

RESULT 14
 LOCUS BM713891 507 bp mRNA linear EST 28-FEB-2002
 DEFINITION UI-E-EJ0-ahq-c-05-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone
 ACCESSION BM713891
 VERSION BM713891.1 GI:19027149
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 507)
 AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@iuii.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 The following repetitive elements were found in this cDNA
 sequence: 443-498, >POLY_ASimple_repeat
 Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers

source

1. 507
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="UI-E-EJ0-abq-c-05-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pTV73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-EJ0 is a subcloned cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pTV73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (GT)₁₈ tail. The sequence tags for this library are: fetal eyes, AGAATCAGACA; lens, CGATTAGCGA; eye anterior segment, AATGCCGAT; optic nerve, CCATTAACTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 184 a 89 c 121 g 112 t 1 others

ORIGIN

Query Match 25.5%; Score 350; DB 14; Length 507;
Best Local Similarity 100.0%; Pred. No. 1.2e-75;
Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 AAATTTTGAGATGCAAAAGCTTTCTGTGAAGACAGCTTCATCTTTTCATTA 1081
|||||
Db 1 AAATTTTGAGATGCAAAAGCTTTCTGTGAAGACAGCTTCATCTTTTCATTA 60

QY 1082 ACACATAGAGAGAGACAGCAATGATATAAAACAGATGATGAGAGAGACCCACTGGA 1141
|||||
Db 61 ACACATAGAGAGAGACAGCAATGATATAAAACAGATGATGAGAGAGACCCACTGGA 120

QY 1142 TCGGCTCAGACAGCTCAGAGCGTGAATGATGAGAGCTGATGGACATCTCCAG 1201
|||||
Db 121 TCGGCTCAGACAGCTCAGAGCGTGAATGATGAGAGCTGATGGACATCTCCAG 180

QY 1202 ACTACAAAAATTGGAAGCTGACACCGGATTAAGTGGGCTATGGCCATGGCCAGAG 1261
|||||
Db 181 ACTACAAAAATTGGAAGCTGACACCGGATTAAGTGGGCTATGGCCATGGCCAGAG 240

QY 1262 AAGACTGTGCTGGTGAATTTATGCTGGGCACTGGAAGATTCGAATGGAAGCTCA 1321
|||||
Db 241 AAGACTGTGCTGGTGAATTTATGCTGGGCACTGGAAGATTCGAATGGAAGCTCA 300

QY 1322 ATAACTTATTTGCGAAAAAGACAGGAGACAGTACTGTCACTGTGATTA 1371
|||||
Db 301 ATAACTTATTTGCGAAAAAGACAGGAGACAGTACTGTCACTGTGATTA 350

RESULT 15
B0771366 808 bp mRNA linear EST 26-JUL-2002
LOCUS B0771366
DEFINITION UI-M-F10-pyu-9-09-0-UI.r1 NIH_BMAP_F10 Mus musculus cDNA clone
IMAGE:5702432.5', mRNA sequence.
ACCESSION B0771366
VERSION B0771366.1 GI:21979842
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 808)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.

TITLE
JOURNAL
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
Seq primer: pyx-5.

FEATURES
source

1. 808
/organism="Mus musculus"
/strain="C57BL/6"
/db.xref="taxon:10090"
/clone="IMAGE:5702432"
/tissue_type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: Brain; Vector: pyx-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pyx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCCACGAC. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT 270 a 158 c 173 g 206 t 1 others

ORIGIN

Query Match 25.1%; Score 343.8; DB 14; Length 808;
Best Local Similarity 85.1%; Pred. No. 5e-74;
Matches 384; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 921 GGCCTCGCAATGAGCCACCCCGACCGAGAGCAATGCGCTGCCCTCACTGGA 980
|||||
Db 1 GGCCTCGCAATGAGCCACCCCGACCGAGAGCAATGCGCTGCCCTCACTGGA 60

QY 981 GAACCTCAGAGCAATGCTACTATTTTCAGTGAAGAAATTTGAGAGTGCANA 1040
|||||
Db 61 GAACCTCAGAGCAATGCTACTATTTTCAGTGAAGAAATTTGAGAGTGCANA 120

QY 1041 GCTTTTCTGGAAGCAATCTTCACTCTTTTTCATTAACACATAGAGAGACAGA 1100
|||||
Db 121 GCTTTTCTGGAAGCAATCTTCCATCTCGTTTTCATTAACACATAGAGAGACAGA 180

QY 1101 ATGAGTAAAAAAGACATGATGAGAGAGAGACCACTGGATGGCTCAGACTCAGA 1160
|||||
Db 181 ATGAGTAAAAAAGACATGATGAGAGAGAGAGCAATGATGGCTCAGACTCAGA 240

QY 1161 GGGTGAATTAATGATGAGTGGCTGATGAGGACATCTCAGACTACAAAAATTTGGAAG 1220
|||||
Db 241 ACAGGAACCGAATGAGTGGCTGATGAGGACATCTCAGACTACAAAAATTTGGAAG 300

QY 1221 TGGACAGCCGATTAATCTGGGCTATGCGCATGGGCGAGAGAGACTGCTGGATTGAT 1280
|||||
Db 301 TGGACAGCCGATTAATCTGGGCTATGCGCATGGGCGAGAGAGACTGCTGGATTGAT 360

QY 1281 TTATGCTGGCGATGAGACGATTTCCATATGTAAGACGCTCAATTAATTTGCAANA 1340
|||||
Db 361 TTACGAGAGACAGATGGAATGACTTCCAGTGTGATGAATCAATTAATTTGATGAGAA 420

QY 1341 AGACAGGAGACAGTACTGTCAATCGCATTA 1371
|| ||||| ||||| ||||| |||||
Db 421 GGAAGGAGGAGCAGTACATCATCATCAATATTA 451

Search completed: March 21, 2003, 08:08:10
Job time : 1811.64 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:02:25 ; Search time 2598.25 seconds
(without alignments)
11492.169 Million cell updates/sec

Title: US-09-763-712A-1_COPY_670_1695

Perfect score: 1026

Sequence: 1 atgcaacaagattgatcgag.....cagtaactgcatcgcattta 1026

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: GenBank:
2: gb-ba:
3: gb-hcg:
4: gb-in:
5: gb-om:
6: gb-ov:
7: gb-pat:
8: gb-ph:
9: gb-pl:
10: gb-pr:
11: gb-ro:
12: gb-sts:
13: gb-sy:
14: gb-un:
15: gb-vi:
16: em-ba:
17: em-fun:
18: em-hum:
19: em-in:
20: em-mu:
21: em-om:
22: em-or:
23: em-ov:
24: em-pat:
25: em-ph:
26: em-pl:
27: em-ro:
28: em-sts:
29: em-un:
30: em-vi:
31: em-hcg-hum:
32: em-hcg-inv:
33: em-hcg-other:
34: em-hcg-mus:
35: em-hcg-pln:
36: em-hcg-rod:
37: em-hcg-mam:
38: em-hcg-vit:
39: em-hcg-hum:
40: em-hcg-mus:
41: em-htgo-other:

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|--------------------|
| 1 | 1026 | 100.0 | 2983 | 9 AB005145 | AB005145 Homo sapi |
| 2 | 1024.4 | 99.8 | 1886 | 9 HSM802985 | AL713657 Homo sapi |
| 3 | 1024.4 | 99.8 | 2005 | 6 AX454442 | AX454442 Sequence |
| 4 | 1024.4 | 99.8 | 2005 | 6 AX490920 | AX490920 Sequence |
| 5 | 1024.4 | 99.8 | 2641 | 6 AX047353 | AX047353 Sequence |
| 6 | 1024.4 | 99.8 | 3058 | 9 AB038518 | AB038518 Homo sapi |
| 7 | 802 | 78.2 | 2637 | 10 AB078434 | AB078434 Mus muscu |
| 8 | 800.4 | 78.0 | 3291 | 10 AB038519 | AB038519 Mus muscu |
| 9 | 615.4 | 60.0 | 4330 | 9 AB052103 | AB052103 Homo sapi |
| 10 | 488.8 | 47.6 | 169088 | 2 AC016128 | AC016128 Homo sapi |
| 11 | 488.8 | 47.6 | 188439 | 9 AP000915 | AP000915 Homo sapi |
| 12 | 484.6 | 47.2 | 71044 | 2 AC024368 | AC024368 Homo sapi |
| 13 | 466.2 | 45.4 | 178022 | 2 AP001022 | AP001022 Homo sapi |
| 14 | 465.8 | 45.4 | 187635 | 2 AP000900 | AP000900 Homo sapi |
| 15 | 437.4 | 42.6 | 188255 | 2 AP000939 | AP000939 Homo sapi |
| 16 | 388 | 37.8 | 182029 | 2 AC114677 | AC114677 Mus muscu |
| 17 | 388 | 37.8 | 193208 | 2 AC102618 | AC102618 Mus muscu |
| 18 | 378.4 | 36.9 | 130763 | 2 AC112416 | AC112416 Rattus no |
| 19 | 157.6 | 15.4 | 3636 | 9 AB007829 | AB007829 Homo sapi |
| 20 | 157.6 | 15.4 | 3685 | 6 E32511 | E32511 Scavenger r |
| 21 | 157.6 | 14.7 | 3810 | 6 E32509 | E32509 Scavenger r |
| 22 | 150.6 | 14.7 | 2215 | 10 BC026446 | BC026446 Mus muscu |
| 23 | 146.8 | 14.3 | 178022 | 2 AP000900 | AP000900 Homo sapi |
| 24 | 146.8 | 14.3 | 187635 | 2 AP001022 | AP001022 Homo sapi |
| 25 | 129 | 12.6 | 6156 | 10 AB009993 | AB009993 Mus muscu |
| 26 | 129 | 12.6 | 188255 | 2 AP000939 | AP000939 Homo sapi |
| 27 | 128.2 | 12.5 | 2193 | 3 PALCOLIA | M25282 P.Lividus a |
| 28 | 125.8 | 12.3 | 6076 | 10 AF272661 | AF272661 Rattus no |
| 29 | 124.2 | 12.1 | 1041 | 4 AF138883 | AF138883 Bos tauru |
| 30 | 124.2 | 12.1 | 2053 | 5 AB008374 | AB008374 Oncohyne |
| 31 | 124.2 | 12.1 | 4502 | 5 AB052836 | AB052836 Oncohyne |
| 32 | 122.6 | 11.9 | 4995 | 10 MMU16789 | U16789 Mus musculu |
| 33 | 121.6 | 11.9 | 810 | 6 AR014116 | AR014116 Sequence |
| 34 | 121.4 | 11.8 | 4628 | 4 AB008683 | AB008683 Bos tauru |
| 35 | 121.4 | 11.8 | 5676 | 6 E07265 | E07265 cDNA encodi |
| 36 | 121.4 | 11.8 | 5676 | 9 HUMCALV | D90279 Human RNA |
| 37 | 121.4 | 11.8 | 7138 | 9 HUMPAIV | M76729 Human pro-a |
| 38 | 121.4 | 11.8 | 4581 | 4 EC62528 | U63528 Equus cabal |
| 39 | 119.8 | 11.7 | 2010 | 9 HUMA2XICOL | L16987 Human alpha |
| 40 | 119.4 | 11.6 | 5551 | 10 AF272662 | AF272662 Rattus no |
| 41 | 118.4 | 11.5 | 6109 | 6 AX239611 | AX239611 Sequence |
| 42 | 118.4 | 11.5 | 6109 | 10 AF176645 | AF176645 Mus muscu |
| 43 | 118 | 11.5 | 756 | 6 AR014110 | AR014110 Sequence |
| 44 | 118 | 11.5 | 6114 | 10 CRUPAIV | M76730 Chinese ham |
| 45 | 118 | 11.5 | 130763 | 2 AC112416 | AC112416 Rattus no |

ALIGNMENTS

```

RESULT 1
AB005145          2983 bp  mRNA  linear  PRI 21-NOV-2001
LOCUS      Homo sapiens CL-P1 mRNA for collectin placenta 1, complete cds.
DEFINITION
AB005145
ACCESSION
VERSION    AB005145.1 GI:17026100
KEYWORDS
SOURCE
ORGANISM   Homo sapiens female tissue_1lb:placenta cDNA to mRNA.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS    Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H.,
            Sakai,Y., Fukuooh,A., Sakamoto,T., Itabe,H., Suzutani,T.,
            Ogasawara,M., Yoshida,I. and Wakamiya,N.

```

TITLE The membrane-type collectin CL-P1 is a scavenger receptor on
JOURNAL J. Biol. Chem. 276 (47), 44222-44228 (2001)
MEDLINE 21570232
REFERENCE 2 (bases 1 to 2983)
AUTHORS Ohtani, K.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1997) Katsuki Ohtani, Asahikawa Medical College,
Department of Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa,
Hokkaido 078-8510, Japan (E-mail: ohtani@asahikawa-med.ac.jp,
Tel: +81-166-68-2393, Fax: +81-166-68-2399)
FEATURES
source
1..2983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="female"
/tissue_lib="placenta"
1..2983
/gene="CL-P1"
71..2299
/gene="CL-P1"
/codon_start=1
/product="collectin placenta 1"
/protein_id="BAB72147.1"
/db_xref="GI:17026101"
/translation="MKDVEEENEVVSFGYKRFGLQEGTQCTCKNNMALKESIIILY
ILCALITVAIIIGYKVEKMDNVGTGEMTSRQYDCKLTAEESDLKGGDTGKKA
STNSLSTERSDILDLROOLREITETSKNKDLKLOAGSLVADLRGSLKETTENN
SFLITVTKTLQAGNYVNLLODPTSVLQONONOMYSHNVIMNLNLQVOORN
LITNLDORSYDTSQAIQRIKNDPQNIQOYFLQAKKTIDLKEKVSLOTLLAANSALA
KANPDLIEDNSQLNSFTGOMENTITTSQANEQNLKDLQDLKDAENRTAIFNQLIE
RQLEFETDITVNIISNISTYAHRLTLISNEMVRYCTPTLTKHNDLISLNNLTANI
LIDVSLRMQODLMSRLDTEYVANSIVMEKMLVDSKGOLIKNFTIIQGGPGPGP
RDRSGSQGPGPGPGSGDPSGPGKEGILGAPGPGEGSGSGSGSPK
GSGSGSGKPGPGSGDPSGPGKEGILGAPGPGEGSGSGSGSPK
LBOVPMSPKPGPRPGSGAVPLALONEPRPAPEDNCGCPHMKNPDKCTYSYE
KEIFEDAKLFECDKSSHLVINTREDOQNIKQWGRSHWIGLTDSEEDNEKMWLDG
TSPDYKNMAGQPDNMNGHGPEDCAGLIVAGQWMDFOCEDVNNPICEKRDYTVLSS
AL"

BASE COUNT 914 a 707 c 703 g 659 t
ORIGIN

Query Match 100.0%; Score 1026; DB 9; Length 2983;
Best Local Similarity 100.0%; Pred. No. 3.8e-226;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAACAAGATTGATGAGGTGAGGTGATGACACTGAATAGCAACTATCAGTATT 60
DB 1271 ATGCACACAGATTGATGAGGTGAGGTGATGACACTGAATAGCAACTATCAGTATT 1330
QY 61 ATGGAAGAATAAGAGCTAGTAGACTCCAGCATGTCAGCTCATCAAGATTTTACATA 120
DB 1331 ATGGAAGAATAAGAGCTAGTAGACTCCAGCATGTCAGCTCATCAAGATTTTACATA 1390
QY 121 CTACAAGTCCACCGGCGCCAGGCGTCCAAGAGGTGACAGAGATCCCGAGCCCT 180
DB 1391 CTACAAGTCCACCGGCGCCAGGCGTCCAAGAGGTGACAGAGATCCCGAGCCCT 1450
QY 181 GGGCCAACTGCAACAAGAGGACAGAAAGAGAGAGAGGAGCTGGACACCTGGCCCT 240
DB 1451 GGGCCAACTGCAACAAGAGGACAGAAAGAGAGAGAGGAGCTGGACACCTGGCCCT 1510
QY 241 GCGGGTGAGAGAGCCCAATTGAGACAGCTGTCCCGCGAGAGCGTGGCGCAAGA 300
DB 1511 GCGGGTGAGAGAGCCCAATTGAGACAGCTGTCCCGCGAGAGCGTGGCGCAAGA 1570
QY 301 TCTAAGGCTCCCAAGGCGCCCAAGGCTCCCGTGGTCCCTGGGAAAGCCCGCCCTCAG 360
DB 1571 TCTAAGGCTCCCAAGGCGCCCAAGGCTCCCGTGGTCCCTGGGAAAGCCCGCCCTCAG 1630
QY 361 GGGCCAGTGGGAGACCGGCGCCCGGCGCGACAGGCAAGAGGAGACTCCCGGCGCT 420
DB 1631 GGGCCAGTGGGAGACCGGCGCCCGGCGCGACAGGCAAGAGGAGACTCCCGGCGCT 1690

QY 421 CAGGCGCCCTCTGCTGCTTCCAGGAGACTTCAGGGACCGCTTGGGGAGCGCTGGCTGGA 480
DB 1691 CAGGCGCCCTCTGCTGCTTCCAGGAGACTTCAGGGACCGCTTGGGGAGCGCTGGCTGGA 1750
QY 481 CCTCGGGAGCTGCCAGGCTTGCCTGGGGTACAGGACATGCGAGCCCAAGGCGCCCGCC 540
DB 1751 CCTCGGGAGCTGCCAGGCTTGCCTGGGGTACAGGACATGCGAGCCCAAGGCGCCCGCC 1810
QY 541 GGGCCCTCTGCGCCATCAGAGAGGCGGTGTCGCCCTGGCCCTGAGATGAGCCACCCCG 600
DB 1811 GGGCCCTCTGCGCCATCAGAGAGGCGGTGTCGCCCTGGCCCTGAGATGAGCCACCCCG 1870
QY 601 GCACCGGAGCAATGAGCTCCCGGCTCAGTGAAGAACTTCCAGACAAATGCTACTAT 660
DB 1871 GCACCGGAGCAATGAGCTCCCGGCTCAGTGAAGAACTTCCAGACAAATGCTACTAT 1930
QY 661 TTTTCACTGAGAAAGAAATTTTGAAGATGCAAAAGCTTTCTGTGACAGCAAGTCTTCA 720
DB 1931 TTTTCACTGAGAAAGAAATTTTGAAGATGCAAAAGCTTTCTGTGACAGCAAGTCTTCA 1990
QY 721 CATCTGTTTCAATTAACACTAGAGAGACACCAATGATTAATAAACAAGATGTAGG 780
DB 1991 CATCTGTTTCAATTAACACTAGAGAGACACCAATGATTAATAAACAAGATGTAGG 2050
QY 781 AGAGAGAGCCACATGATCGGCTCAGAGACTCAGAGCGTGAATGATGAAGTGGCTG 840
DB 2051 AGAGAGAGCCACATGATCGGCTCAGAGACTCAGAGCGTGAATGATGAAGTGGCTG 2110
QY 841 GATGGGACATCTCCAGACTACAAAAATTTGAAGCTGAGACGCGGATTAATGAGGCTAT 900
DB 2111 GATGGGACATCTCCAGACTACAAAAATTTGAAGCTGAGACGCGGATTAATGAGGCTAT 2170
QY 901 GGGCATGGGCGGAGAGAAAGCTGCGTGGTATTTATCTGGGCGAGTGAACGATTC 960
DB 2171 GGGCATGGGCGGAGAGAAAGCTGCGTGGTATTTATCTGGGCGAGTGAACGATTC 2230
QY 961 CAATGTGAAGAGCTCAATACTTCAATTTGGGAAAAGACAGGAGACAGTACTGATCT 1020
DB 2231 CAATGTGAAGAGCTCAATACTTCAATTTGGGAAAAGACAGGAGACAGTACTGATCT 2290
QY 1021 GCATTA 1026
DB 2291 GCATTA 2296

RESULT 2
LOCUS HSM802985
DEFINITION Homo sapiens mRNA; cDNA DKFZp547G1215 (from clone DKFZp547G1215).
ACCESSION AL713657
VERSION AL713657.1 GI:19584339
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1886)
AUTHORS Koehler, K., Beyer, A., Mewes, H.W., Well, B. and Wiemann, S.
TITLE Submitted (12-MAR-2002) MIPs, Am klopfersplitz 18a, D-82152
JOURNAL Martinried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFZp547G1215) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cdna/>.
FEATURES
Location/Qualifiers

| source | 1. .1886 |
|--------------|---|
| gene | /organism="Homo sapiens" |
| CDS | /db_xref="taxon:9606" |
| | /map="18p11.3" |
| | /clone="DKFzp547G1215" |
| | /tissue_type="brain" |
| | /clone_id="547 (synonym: hfbri). Vector pSport1; host DH10b; sites NotI + SalI" |
| | /dev_stage="fetal" |
| | 1. .1886 |
| | /gene="DKFzp547G1215" |
| | <1. .1256 |
| | /gene="DKFzp547G1215" |
| | /note="similarity to asialoglycoprotein receptor (Mus musculus" |
| | /codon_start=3 |
| | /product="hypothetical protein" |
| | /protein_id="CAD28466.1" |
| | /db_xref="GI:19584340" |
| | /translation="AIKPNQLEERPOLFETDIVNIISNISTAHHLRTLSLNEVRFITCTDITHTHDTLSLTNNLTLANIRLDSSLRMOODLSRLDTEVANSVIMEEMKILVDSKHQILKNFTIILOGPPGPRGPRGDRGSSQGPPTGKKGQKGGKGGPPGAPGGEGLPGAPGPGGSGGSGKSGSGKSGPRGSGSGDPPGPPREKEELPPPGGPPGPGQLQSTVGEPPVPPRGPLPLGLGVGMGPYPGPPGPPGSGAVVPLALQNEPTPLPEDNCSPHMKNNFTDKCYFSEVEKIEFDALFCEDKSSHLVFTNTREDOQMIKQWVGRESNMIQLTOSERENEMKWLDDGTSPPDKNNKAGQPDNMGHGHDGEDCAQLIYAQOMNDFOCEDNNFTICEKDRETVLSAL" |
| polyA_signal | 1838. .1843 |
| polyA_site | /gene="DKFzp547G1215" |
| COUNT | 1869 "DKFzp547G1215" |
| | /gene="DKFzp547G1215" |
| | 444 C 445 G 428 T |

| | | | | |
|-----------------------|----------------|--------------------|----------|-------------|
| Query Match | 99.8% | Score 1024.4 | DB 9 | Length 1886 |
| Best Local Similarity | 99.9% | Pred. No. 8.7e-226 | | |
| Matches 1025 | Conservative 0 | Mismatches 1 | Indels 0 | Gaps 0 |

| | | | |
|----|-----|---|-----|
| QY | 1 | ATGCAACAAGATTTTGATGAGTCGAGGTTAGACATGAGAGTACCACCTATACATGATT | 60 |
| DB | 228 | ATGCAMCAAGATTTTGATGATGAGGTCGAGGTTAGACATGAAAGTACCACTATACATGATT | 287 |
| QY | 61 | ATGGAAGAAATGABAGCTGTGACTCCAAAGCATGTGCTACGCTCATCAAGATTTTACAATA | 120 |
| DB | 288 | ATGGAAGAAATGAACTGTAGACTCCAAAGCATGTGCTACGCTCATCAAAATTTTACAATA | 347 |
| QY | 121 | CTACAAAGTCCACCGGGGGCCCAAGGGGTCCAAAGGTTGACAGAGATATCCAGGAGCCCT | 180 |
| DB | 348 | CTACAAAGTCCACCGGGGGCCCAAGGGGTCCAAAGGTTGACAGAGATATCCAGGAGCCCT | 407 |
| QY | 181 | GGCCCACTGGCAACMAAGGAGACAGAAAGAGAGAAGGGGGAGCCCTGGAGCAACTGGCCCT | 240 |
| DB | 408 | GGCCCACTGGGAACMAAGGAGACAGAAAGAGAGAAGGGGGAGCCCTGGAGCAACTGGCCCT | 467 |
| QY | 241 | GGGGGTGAGAGAGGCCCAATTGGACACAGCTGGTCCCCCGGAGAGCGTGGGGCAAGAA | 300 |
| DB | 468 | GGGGGTGAGAGAGGCCCAATTGGACACAGCTGGTCCCCCGGAGAGCGTGGGGCAAGAA | 527 |
| QY | 301 | TCTAAAGCTCCCAAGGGGCCCAAGGCTCCCGTGGTTCCCTGGGAAACCCGGCCCTAG | 360 |
| DB | 528 | TCTAAAGCTCCCAAGGGGCCCAAGGCTCCCGTGGTTCCCTGGGAAACCCGGCCCTAG | 587 |
| QY | 361 | GGCCCACTGGGAGACCAAGGCCCCCGGGGCCACACAGGCAAAAGAGAGACTCCCGGGCCT | 420 |
| DB | 588 | GGCTCAATGGGAGACCAAGGCCCCCGGGGCCACACAGGCAAAAGAGAGACTCCCGGGCCT | 647 |
| QY | 421 | CAGGCGCTCTCGCTTCAGAGGACTTCAGGACACCGTTGGGGAGCTGGGGGTGCTTGA | 480 |
| DB | 648 | CAGGCGCTCTCTCGCTTCAGAGGACTTCAGGAGCACCGTTGGGGAGCTGGGGGTGCTTGA | 707 |
| QY | 481 | CTGGGGGACAGGCAAGGTTGCTGGGGTTACAGGCAATGCCAGAGCCCAAGAGGCCCCCCC | 540 |
| DB | 708 | CTGGGGGACAGGCAAGGTTGCTGGGGTTACAGGCAATGCCAGAGCCCAAGAGGCCCCCCC | 767 |

| | | | |
|----|------|---|------|
| QY | 541 | GGCCCTCCTGGGCCCACTCAGAGACGGGTGGGCCCTCGGCCCTGCAACAATGAGCAACCCG | 600 |
| Db | 768 | GGCCCTCCTGGGCCCACTCAGAGACGGGTGGGCCCTCGGCCCTGCAACAATGAGCAACCCG | 827 |
| QY | 601 | GCACCGGAGGACAAATGGCTGCCCGCCTCAGTGGAGAACTTCACAGACAATGCTACTAT | 660 |
| Db | 828 | GCACCGGAGGACAAATGGCTGCCCGCCTCAGTGGAGAACTTCACAGACAATGCTACTAT | 887 |
| QY | 661 | TTTTAGATTGGAAGAAATTTTGGAGATGGAAGCTTTTCTGTGAAGACAGCTTCA | 720 |
| Db | 888 | TTTTAGATTGGAAGAAATTTTGGAGATGGAAGCTTTTCTGTGAAGACAGCTTCA | 947 |
| QY | 721 | CATCTTGTTTTCATTAACACTAGAGAGAACAGCAATGGATAAAAAAAACAGATGGTAGGG | 780 |
| Db | 948 | CATCTTGTTTTCATTAACACTAGAGAGAACAGCAATGGATAAAAAAAACAGATGGTAGGG | 1007 |
| QY | 781 | AGAGAGACCACTGGATTTGGGCTCCACACACTCAGAGCGCTGAAAAATGAGTAAGTGGCTG | 840 |
| Db | 1008 | AGAGAGACCACTGGATTTGGGCTCCACACACTCAGAGCGCTGAAAAATGAGTAAGTGGCTG | 1067 |
| QY | 841 | GATGGACATCTCCAGACTACAAAAATTGGAAGCTGGACAGCCGATTACTGGGGTCAT | 900 |
| Db | 1068 | GATGGACATCTCCAGACTACAAAAATTGGAAGCTGGACAGCCGATTACTGGGGTCAT | 1127 |
| QY | 901 | GGCCATTTGGGCCAGGAAGACTGTGCTGGGTGATTATGCTGGGCACTGGAAACGATTTC | 960 |
| Db | 1128 | GGCCATTTGGGCCAGGAAGACTGTGCTGGGTGATTATGCTGGGCACTGGAAACGATTTC | 1187 |
| QY | 961 | CAATTTGGAAGACGTCAATTAACCTCATTTTGGCAAAAAAGACAGGGAACAGTACTGTCATCT | 1020 |
| Db | 1188 | CAATTTGGAAGACGTCAATTAACCTCATTTTGGCAAAAAAGACAGGGAACAGTACTGTCATCT | 1247 |
| QY | 1021 | GCATTA 1026 | |
| Db | 1248 | GCATTA 1253 | |

| | | | | | |
|-----------------------|---|-------------|-----|--------|-----------------|
| AX545442 | RESULT 3 | | | | |
| LOCUS | AX544442 | 2005 bp | DNA | linear | PAT 06-JUL-2002 |
| DEFINITION | Sequence 27 from Patent WO2008284. | | | | |
| ACCESSION | AX544442 | | | | |
| VERSION | AX544442.1 | GI:21713845 | | | |
| KEYWORDS | | | | | |
| SOURCE | human. | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| AUTHORS | 1 Baker,K.P., Ferrara,N., Gerber,H., Gerritsen,M.E., Goddard,A., Godowski,P.J., Gurey,A.L., Hillan,K.J., Marsters,S.A., Pan,J., Paoni,N.F., Stephan,J.P., Watanabe,C.K., Williams,P.M., Wood,W.I. and Ye,W. | | | | |
| TITLE | Compositions and methods for the diagnosis and treatment of disorders involving anglogenesis | | | | |
| JOURNAL | Patent: WO 0208284-A 27 31-JAN-2002; Genentech, Inc. (US) ; Baker, Kevin P. (US) ; Ferrara, Napoleone (US) ; Gerber, Hanspeter (US) ; Gerritsen, Mary E. (US) ; Goddard, Audrey (US) ; Godowski, Paul J. (US) ; Gurey, Austin L. (US) ; Hillan, Kenneth J. (US) ; Marsters, Scott A. (US) ; Pan, James (US) ; Paoni, Nicholas F. (US) ; Stephan, Jean-Philippe F. (US) ; Watanabe, Colin K. (US) ; Williams, P. Mickey (US) ; Wood, William I. (US) | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..2005 /organism="Homo sapiens" /db_xref="taxon:9606" | | | | |
| BASE COUNT | 606 a 493 c 491 g 415 t | | | | |
| ORIGIN | | | | | |
| Query Match | 99.8% ; Score 1024.4 ; DB 6 ; Length 2005 ; | | | | |
| Best Local Similarity | 99.9% ; Pred. No. 8.7e-226 ; | | | | |

| | Matches | 1025: | Conservative | 0: | Mismatches | 1: | Indels | 0: | Gaps |
|----|---------|---|--------------|----|------------|----|--------|----|------|
| Qy | 1 | ATGCACAAGAGTTTATGATGAGTCGAGGTTAGACACTGAAAGTATGCCACTTATCATGTGATT | 60 | | | | | | |
| Db | 478 | ATGCACAAGAGTTTATGATGAGTCGAGGTTAGACACTGAAAGTATGCCACTTATCATGTGATT | 537 | | | | | | |
| Qy | 61 | ATGGAAGAATGAACTAGTAGACTCCAAAGCATYGGTCAGCTCATCAAGAAATTTTACAAAT | 120 | | | | | | |
| Db | 538 | ATGGAAGAATGAACTAGTAGACTCCAAAGCATYGGTCAGCTCATCAAGAAATTTTACAAAT | 597 | | | | | | |
| Qy | 121 | CTACAGGTCCACCGGGCCCCAGGGGTCCAGAGGTGACAGAGATCCAGGGACCCCT | 180 | | | | | | |
| Db | 598 | CTACAGGTCCACCGGGCCCCAGGGGTCCAGAGGTGACAGAGATCCAGGGACCCCT | 657 | | | | | | |
| Qy | 181 | GGCCCACTGGCAACAAGGACAGAAAGGAGAAAGGGGAGACCTGGACCTGGCCCT | 240 | | | | | | |
| Db | 658 | GGCCCACTGGCAACAAGGAGACGAAAGAGAAAGGGGAGACCTGGACCTGGCCCT | 717 | | | | | | |
| Qy | 241 | GGGGGTGAGAGAGGCCCAATTTGACAGCTGTCCCCCGGAGAGCGTGGCGCAAGGA | 300 | | | | | | |
| Db | 718 | GGGGGTGAGAGAGGCCCAATTTGACAGCTGTCCCCCGGAGAGCGTGGCGCAAGGA | 777 | | | | | | |
| Qy | 301 | TCTAAGGCTCCAGGGGCCCAAGGCTCCGGTGTCCCTCGGGAAGGCCGGCCCTCAG | 360 | | | | | | |
| Db | 778 | TCTAAGGCTCCAGGGGCCCAAGGCTCCGGTGTCCCTCGGGAAGGCCGGCCCTCAG | 837 | | | | | | |
| Qy | 361 | GGCCCCAGTGGGGACCCACAGGCCCCCCGGGCCCAACAGGCAAGAGAGGATCCCGGGCCT | 420 | | | | | | |
| Db | 838 | GGCCCCAGTGGGGACCCACAGGCCCCCCGGGGCCCAACAGGCAAGAGGATCCCGGGCCT | 897 | | | | | | |
| Qy | 421 | CAGGGCCCTCTGGCTTCAGAGGACTTCAGGGCACCGTGGGGAGCTTGGGTGCTTGA | 480 | | | | | | |
| Db | 898 | CAGGGCCCTCTGGCTTCAGAGGACTTCAGGGCACCGTGGGGAGCTTGGGTGCTTGA | 957 | | | | | | |
| Qy | 481 | CTTCGGGAGACTCCAGGCTTGCTGGGGTACAGAGCATCCAGGCCCCCAAGGGCCCCC | 540 | | | | | | |
| Db | 958 | CTTCGGGAGACTCCAGGCTTGCTGGGGTACAGAGCATCCAGGCCCCCAAGGGCCCCC | 1017 | | | | | | |
| Qy | 541 | GGCCCTCTGGCCCATCAGAGAGCGGTGCTGCCCTTGCCCTGACAGANTGAGCCAACCCG | 600 | | | | | | |
| Db | 1018 | GGCCCTCTGGCCCATCAGAGAGCGGTGCTGCCCTTGCCCTGACAGANTGAGCCAACCCG | 1077 | | | | | | |
| Qy | 601 | GCACCGGAGAGCAATGGCTGCCCGCTCACCTGGAAGAACTTCACAGACAAATGCTACTAT | 660 | | | | | | |
| Db | 1078 | GCACCGGAGAGCAATGGCTGCCCGCTCACCTGGAAGAACTTCACAGACAAATGCTACTAT | 1137 | | | | | | |
| Qy | 661 | TTTTCAGTTGAGAAAGAAATTTTGTAGAGTGCAGAAAGCTTTTCTGTGAAGACAAGTCTTCA | 720 | | | | | | |
| Db | 1138 | TTTTCAGTTGAGAAAGAAATTTTGTAGAGTGCAGAAAGCTTTTCTGTGAAGACAAGTCTTCA | 1197 | | | | | | |
| Qy | 721 | CATCTGTTTTTCATAAACATGAGAGAAACGCAATGGATATAAAAAACAGATGGTAGGG | 780 | | | | | | |
| Db | 1198 | CATCTGTTTTTCATAAACATGAGAGAAACGCAATGGATATAAAAAACAGATGGTAGGG | 1257 | | | | | | |
| Qy | 781 | AGAGAGAGCACTGGATGGGCTTCACAGACTCAGAGCTGAGAGCTGAAATATGAAATGAGTGGCTG | 840 | | | | | | |
| Db | 1258 | AGAGAGAGCACTGGATGGGCTTCACAGACTCAGAGCTGAGAGCTGAAATATGAAATGAGTGGCTG | 1317 | | | | | | |
| Qy | 841 | GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGAGTAACTTGGGGTCAAT | 900 | | | | | | |
| Db | 1318 | GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGAGTAACTTGGGGTCAAT | 1377 | | | | | | |
| Qy | 901 | GGCCATYGGGCCAGGAGAGACTGTCTGGGTGATTTATGCTGGGGCAGTGGAAAGCATTTTC | 960 | | | | | | |
| Db | 1378 | GGCCATYGGGCCAGGAGAGACTGTCTGGGTGATTTATGCTGGGGCAGTGGAAAGCATTTTC | 1437 | | | | | | |
| Qy | 961 | CAATGTGAAGAGCTCAATTAATCTTCTTTGCCAAAAAGACAGGAGAGACAGTACTGTATCT | 1020 | | | | | | |
| Db | 1438 | CAATGTGAAGAGCTCAATTAATCTTCTTTGCCAAAAAGACAGGAGAGACAGTACTGTATCT | 1497 | | | | | | |
| Qy | 1021 | GCATTA 1026 | | | | | | | |
| Db | 1498 | GCATTA 1503 | | | | | | | |

| AX490920 | AX90920 | Sequence 27 from Patent WO0200690. | 2005 bp | DNA | Linear | PAT 16-AUG-2002 |
|----------------------------|--|--|---------|-----|--------|-----------------|
| LOCUS | AX490920 | | | | | |
| DEFINITION | Sequence 27 from Patent WO0200690. | | | | | |
| ACCESSION | AX490920 | | | | | |
| VERSION | AX490920.1 | GI:22323797 | | | | |
| KEYWORDS | | | | | | |
| SOURCE | human. | | | | | |
| ORGANISM | Homo sapiens | | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | | |
| AUTHORS | Baker,K.P., Ferrara,N., Gerber,H., Gertliss,M.E., Goddard,A., Godowski,P.J., Gurely,A.L., Hillian,K.J., Marsters,S.A., Pan,J., Paoni,N.F., Stephan,J.P., Watanabe,C.K., Williams,P.M., Wood,W.I. and Ye,W. | | | | | |
| TITLE | Compositions and methods for the diagnosis and treatment of disorders involving angiogenesis | | | | | |
| JOURNAL | Patent: WO 0200690-A 27 03-JAN-2002; | | | | | |
| FEATURES | Genetech, Inc. (US) | | | | | |
| Source | Location/Qualifiers | | | | | |
| BASE COUNT | 1..2005 | | | | | |
| ORIGIN | "/organism="Homo sapiens" /db_xref="taxon:9606" | | | | | |
| Query Match | Best Local Similarity | 99.8%; Score 1024.4; DB 6; Length 2005; | | | | |
| Matches 1025; Conservative | 0; Mismatches | 1; Indels | 0; Gaps | 0; | | |
| QY | 1 | ATGCACAAAGATTTATGATGAGTCGAGGTGTAGACATCGAAGTAGCCACTTATCAGTATT | 60 | | | |
| DB | 478 | ATGCACAAAGATTTATGATGAGTCGAGGTGTAGACATCGAAGTAGCCACTTATCAGTATT | 537 | | | |
| QY | 61 | ATGGAAGAATGAAGCTAGTAGATCCCAAGCTGAGCTTCATCAAGATTTTACAATA | 120 | | | |
| DB | 538 | ATGGAAGAATGAAGCTAGTAGATCCCAAGCTGAGCTTCATCAAGATTTTACAATA | 597 | | | |
| QY | 121 | CTACAAGTTCACCGGGGCCCGAGGGGTCCAGAGGTGACAGAGATCCAGGAGCCCT | 180 | | | |
| DB | 598 | CTACAAGTTCACCGGGGCCCGAGGGGTCCAGAGGTGACAGAGATCCAGGAGCCCT | 657 | | | |
| QY | 181 | GGCCCACTGGCAACAAGGACAGAAAAGAGAGAGGGGAGCCCTGAGACACTTGACCT | 240 | | | |
| DB | 658 | GGCCCACTGGCAACAAGGAGACAGAAAAGAGAGGGGAGCCCTGAGACACTTGACCT | 717 | | | |
| QY | 241 | GGGGTGAAGAGAGGCCAATTTGACCTGTGCCCCCGGAGAGCGTGGCGGCAAGGA | 300 | | | |
| DB | 718 | GGGGTGAAGAGAGGCCAATTTGACCTGTGCCCCCGGAGAGCGTGGCGGCAAGGA | 777 | | | |
| QY | 301 | TCTAAAGCTCCCAAGGGCCCAAGGCTCCGTGGTCCCTCTGGGAAGCGCGGCTCTAG | 360 | | | |
| DB | 778 | TCTAAAGCTCCCAAGGGCCCAAGGCTCCGTGGTCCCTCTGGGAAGCGCGGCTCTAG | 837 | | | |
| QY | 361 | GGCCCACTGGGAGACCAAGGCCCCCGGAGCCACCAAGGCAAGAGAGGATCCCGGACCT | 420 | | | |
| DB | 838 | GGCCCACTGGGAGACCAAGGCCCCCGGAGCCACCAAGGCAAGAGAGGATCCCGGACCT | 897 | | | |
| QY | 421 | CAGGCGCTCTGCTGCTTCCAGGACTTCAAGGACACGTTGGGAGCTGGGGTGCCTGA | 480 | | | |
| DB | 898 | CAGGCGCTCTGCTGCTTCCAGGACTTCAAGGACACGTTGGGAGCTGGGGTGCCTGA | 957 | | | |
| QY | 481 | CCTGGGGAATGCAAGGCTTGGCTTGGGATACAGGCAATGCAAGGCCCCCAAGGCCCCC | 540 | | | |
| DB | 958 | CCTGGGGAATGCAAGGCTTGGCTTGGGATACAGGCAATGCAAGGCCCCCAAGGCCCCC | 1017 | | | |
| QY | 541 | GGCCCTCTGGGCAATGAGAGCGGTGTGCCCCGGGCTGCAAGATGAGCAACCCCG | 600 | | | |
| DB | 1018 | GGCCCTCTGGGCAATGAGAGCGGTGTGCCCCGGGCTGCAAGATGAGCAACCCCG | 1077 | | | |

QY 601 GCACGGAGACATGCTGCCCGCTCACTGGAGAACTTCACAGACAAATGCTACTAT 660
|||||
Db 1078 GCACGGAGAGCAATGCTGCCCGCTCACTGGAGAACTTCACAGACAAATGCTACTAT 1137
QY 661 TTTTCAGTTGAGAAATTTTGTGAGATGCAAACTTTTCGTGAAGCAAGTCTTCA 720
|||||
Db 1138 TTTTCAGTTGAGAAATTTTGTGAGATGCAAACTTTTCGTGAAGCAAGTCTTCA 1197
QY 721 CATCTGTTTTCATAAACACTAGAGAGAAAGCAATGATATAAAACAGATGTAGGG 780
|||||
Db 1198 CATCTGTTTTCATAAACACTAGAGAGAAAGCAATGATATAAAACAGATGTAGGG 1257
QY 781 AGAGAGAGCCACTGATGAGCGCTCACAGACTCAGAGCGTGAATAATGGAAGTGGCTG 840
|||||
Db 1258 AGAGAGAGCCACTGATGAGCGCTCACAGACTCAGAGCGTGAATAATGGAAGTGGCTG 1317
QY 841 GATGGAGACATCTCCAGACATCAAAAATTTGAAAGCTGACAGCGGATTAAGTGGGTCA 900
|||||
Db 1318 GATGGAGACATCTCCAGACATCAAAAATTTGAAAGCTGACAGCGGATTAAGTGGGTCA 1377
QY 901 GGCCATGGGCCAGAGAACTGTGCTGGGTTGATTTATGCTGGGAGTGAAGCAATTC 960
|||||
Db 1378 GGCCATGGGCCAGAGAACTGTGCTGGGTTGATTTATGCTGGGAGTGAAGCAATTC 1437
QY 961 CATGTGAAGAGCTCAATTAATTCATTTGCGAAAAAGACAGGAGAGACAGTACTGATCT 1020
|||||
Db 1438 CATGTGAAGAGCTCAATTAATTCATTTGCGAAAAAGACAGGAGAGACAGTACTGATCT 1497
QY 1021 GCATTA 1026
|||||
Db 1498 GCATTA 1503

RESULT 5
AX047353 2641 bp DNA linear PAT 15-DEC-2000
LOCUS AX047353
DEFINITION Sequence 39 from Patent WO0068380.
ACCESSION AX047353
VERSION AX047353.1 GI:11876599
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Bandman, O., Hillman, J.L., Tang, Y.T., Lal, P., Yue, H., Baughn, M.R.,
Lu, D.A. and Azimzal, Y.
TITLE Extracellular matrix and adhesion-associated proteins
JOURNAL Patent: WO 0068380-A 39 16-NOV-2000;
Incyte Genomics, Inc. (US)
FEATURES
Source Location/Qualifiers
1..2641
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 823 a 628 c 608 g 582 t
ORIGIN

Query Match 99.8%; Score 1024.4; DB 6; Length 2641;
Best Local Similarity 99.9%; Pred. No. 8.8e-226;
Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAACAGATTTGATGAGTGCAGGTAGACACTGAGTACGCCAATTAATCAGTATT 60
|||||
Db 912 ATGCAACAGATTTGATGAGTGCAGGTAGACACTGAGTACGCCAATTAATCAGTATT 971
QY 61 ATGGAAGAATGAGTACTAGACTCCAGCATGTGTCAGCTCATCAAGATTTTACAATA 120
|||||
Db 972 ATGGAAGAATGAGTACTAGACTCCAGCATGTGTCAGCTCATCAAGATTTTACAATA 1031
QY 121 CTACAAAGTCTCCACGGGGCCCAAGAGGTCCAGAGAGATCCAGGAGCCCCCT 180
|||||
Db 1032 CTACAAAGTCTCCACGGGGCCCAAGAGGTCCAGAGAGATCCAGGAGCCCCCT 1091

QY 181 GCCCAACTGGCAACAAGGACAGAAAGAGAGAGAGGGAGGAGCTTGACACCTGGCCCT 240
|||||
Db 1092 GCCCAACTGGCAACAAGGACAGAAAGAGAGAGAGGGAGGAGCTTGACACCTGGCCCT 1151
QY 241 GCGGGTGAAGAGAGCCCAATTTGACCAAGTGTGCCCGGAGAGAGCTGGCGCAAGA 300
|||||
Db 1152 GCGGGTGAAGAGAGCCCAATTTGACCAAGTGTGCCCGGAGAGAGCTGGCGCAAGA 1211
QY 301 TCTAAAGGCTCCCAAGGCCCCCAAGGCTCCCGGTGTCCCTGGGAAAGCCCGGCTCAG 360
|||||
Db 1212 TCTAAAGGCTCCCAAGGCCCCCAAGGCTCCCGGTGTCCCTGGGAAAGCCCGGCTCAG 1271
QY 361 GCCCCAGTGGGAGACCCAGGCCCCCGGGGCCACAGGCAAAAGAGGACTCCCGGCTC 420
|||||
Db 1272 GCCCCAGTGGGAGACCCAGGCCCCCGGGGCCACAGGCAAAAGAGGACTCCCGGCTC 1331
QY 421 CAGGGCCCTCCTGGCTTCACAGGACCTTCAGGGACCCGTTGGGAGGCTGGGGTGCCTGA 480
|||||
Db 1332 CAGGGCCCTCCTGGCTTCACAGGACCTTCAGGGACCCGTTGGGAGGCTGGGGTGCCTGA 1391
QY 481 CCTCGGGAGACTGCCAGGCTTGCTGGGGTACAGGCAATGCCAGGCCCCCAAGGCTCC 540
|||||
Db 1392 CCTCGGGAGACTGCCAGGCTTGCTGGGGTACAGGCAATGCCAGGCCCCCAAGGCTCC 1451
QY 541 GGGCTTCCTGGGCCATCAGGAGGCGGTGGTCCCTGGGCCCTGCAGAAATGAGCAACCCG 600
|||||
Db 1452 GGGCTTCCTGGGCCATCAGGAGGCGGTGGTCCCTGGGCCCTGCAGAAATGAGCAACCCG 1511
QY 601 GCACCGGAGACAAATGGCTGCCCGCTCACTGAGAAACTTCAACAGCAATGCTACTAT 660
|||||
Db 1512 GCACCGGAGACAAATGGCTGCCCGCTCACTGAGAAACTTCAACAGCAATGCTACTAT 1571
QY 661 TTTTCAGTTGAGAAATTTTGTGAGATGCAAGCTTTTCTGTGAAGCAAGTCTTCA 720
|||||
Db 1572 TTTTCAGTTGAGAAATTTTGTGAGATGCAAGCTTTTCTGTGAAGCAAGTCTTCA 1631
QY 721 CATCTGTTTTCATAAACACTAGAGAGAAAGCAATGATATAAAACAGATGTAGGG 780
|||||
Db 1632 CATCTGTTTTCATAAACACTAGAGAGAAAGCAATGATATAAAACAGATGTAGGG 1691
QY 781 AGAGAGAGCCACTGATGAGCTGCCCTCACAGACTCAGAGCGTGAATAATGGAAGTGGCTG 840
|||||
Db 1692 AGAGAGAGCCACTGATGAGCTGCCCTCACAGACTCAGAGCGTGAATAATGGAAGTGGCTG 1751
QY 841 GATGGAGACATCTCCAGACATCAAAAATTTGAAAGCTGACAGCGGATTAAGTGGGTCA 900
|||||
Db 1752 GATGGAGACATCTCCAGACATCAAAAATTTGAAAGCTGACAGCGGATTAAGTGGGTCA 1811
QY 901 GGCCATGGGCCAGAGAAAGCTGTGCTGGGTTGATTTATGCTGGGAGTGAAGCAATTC 960
|||||
Db 1812 GGCCATGGGCCAGAGAAAGCTGTGCTGGGTTGATTTATGCTGGGAGTGAAGCAATTC 1871
QY 961 CAATGTGAAGAGCTCAATTAATTCATTTGCGAAAAAGACAGGAGACAGTACTGATCT 1020
|||||
Db 1872 CAATGTGAAGAGCTCAATTAATTCATTTGCGAAAAAGACAGGAGACAGTACTGATCT 1931
QY 1021 GCATTA 1026
|||||
Db 1932 GCATTA 1937

RESULT 6
AB038518 3058 bp mRNA linear PRI 08-MAR-2001
LOCUS AB038518
DEFINITION Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
type I, complete cds.
ACCESSION AB038518
VERSION AB038518.1 GI:13365514
KEYWORDS
SOURCE Homo sapiens tissue_lhb:Placenta cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

| | |
|----------------|--|
| TITLE | Molecular cloning and functional characterization of a human |
| AUTHORS | Nakamura, K., Funakoshi, H., Miyamoto, K., Tokunaga, F. and Nakamura, I. |

TITLE

JOURNAL
OF
BIOCHEM. BIOPHYS. RES. COMMUN. 280 (4), 1028-1035 (2001)

JOURNAL
NEXT TIME

| REFERENCE | 2 | (bases 1 to 3058) |
|-----------|----------|-------------------|
| MEDLINE | 21092718 | |

REFERENCE

| | |
|---------|-------------------------------|
| AUTHORS | Nakamura, K. and Nakamura, T. |
| TITLE | Direct Submission |

TITLE

JOURNAL Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical School, Division of Biochemistry, Biomedical Research Center: 2-2


```

/product="scavenger receptor with C-type lectin"
/protein_id="BAB82497.1"
/db_xref="GI:18146952"
/translation="MKDPAEEBEVPSFGFRFGIHEGTQCTKCIINMALKEFISVILY
ILCALITVITVILGKVVKEKMDNVSQDGTETSKNDTLKLAQSDALVDSRSLKEDGAGKAL
SNSELSFTRSDIIDLROOLITETKTSKNDTLKLAQSDALVDSRSLKEDGAGKAL
SPILITVVKTLQAIYNGYTNLQODITVNOGINSOSVYIMNLINLNUVOQRN
LISNLQSDVDTSLAIORIKNDPQNLQOVFLQAKDIDWLKEKVSLOTLANNSALA
KANNDTLEDMSQLSFTGQMDNITTTISQANDQSLKLDLDRKDENRVAFSOLEE
RPOVETDVIINISYTAHRLRTLSMNDVTTCTDITLRTDIDLTLNNTLVNI
RDSISLRQODMRSKLDTEVANSVVEEMKLVDSKGLINKNTILQGPGRGP
KDRGSQGPPTGKNGKQKGEKPEPGEKRGITGVPGSEVDSKSGSGSPK
GSRGSPKRGPGSPGDRGPPGPKDGLPGQGPGRGLOGTGSEVPVPRGLPG
LPVPGMPKPPKPPPPPSGMBELADINEPTVASEVGCPPHMKNTFDKCTYSLE
KEILDALKFECDKSHLVFINSREQDQWIKHVTYGRSHMIGLADSEQSEMKWLDG
SPVDYKMKMAGQPDNMWGSQHGPEDCAGLIVAGWNDFQCDINNEICKEFEAVPSS
IL"
polyA_signal 3246..3251
BASE COUNT 1013 a 757 c 759 g 762 t
ORIGIN
Query Match 78.0%; Score 800.4; DB 10; Length 3291;
Best local similarity 86.3%; Pred. No. 4,2e-174;
Matches 885; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 1 ATGCAACAAGATTGATGAGTGAGGTTAGACACGAAAGTACCAACTTATCAGTGATT 60
DB 1277 ATGCAACAAGATTGATGAGTGAGGTTAGACACGAAAGTACCAACTTATCAGTGATT 1336
QY 61 ATGGAAGAATGAGTACTAGTACTCCAAAGCATGTCAGTCTATCAAGATTTTACAATA 120
DB 1337 ATGGAAGAATGAGTACTAGTACTCCAAAGCATGTCAGTCTATCAAGATTTTACAATA 1396
QY 121 CTACAGATTCACCGGGCCCAAGGGTCCAAAGGTGACAGAGATCCCAAGGACCCCT 180
DB 1397 CTACAGATTCACCGGGCCCAAGGGTCCAAAGGTGACAGAGATCCCAAGGACCCCT 1456
QY 181 GGCCTCACTGGCAACAAGGAGACAGAAAGAGAGAGAGAGGAGGAGCCTTGACCTG 240
DB 1457 GGCCTCACTGGCAACAAGGAGACAGAAAGAGAGAGAGAGGAGGAGCCTTGACCTG 1516
QY 241 GCGGGTGAAGAGAGCCCAATTTGGACAGTGTCCCGGGAGAGCGTGGCGGCAAGGA 300
DB 1517 GCGGGTGAAGAGAGCCCAATTTGGACAGTGTCCCGGGAGAGCGTGGCGGCAAGGA 1576
QY 301 TCTAAGAGTCCCGAGGGCCCAAGGCTCCCGGTTCCCGTGGGAAGCCCGGCTCTG 360
DB 1577 TCTAAGAGTCCCGAGGGCTCCCAAGGATCTCGTGGTCCCGGAGGAGCCTTGCCCTCA 1636
QY 361 GGCCTCACTGGGAGACCCAGGCCCCCGGGCCCAAGGAGAGAGGAGTCCCGGCGCT 420
DB 1637 GGCCTCACTGGGAGACCCAGGCCCCCGGGCCCAAGGAGAGGAGTCCCGGCGCT 1696
QY 421 CAGGGCCCTCTGCTTCCAGGAGCTTCAAGGAGTGGGAGGCTTGGGGTGGCTGGGA 480
DB 1697 CAGGGCCCTCTGCTTCCAGGAGCTTCAAGGAGTGGGAGGCTTGGGGTGGCTGGGA 1756
QY 481 CCTCGGGAGTGGCAGGCTTGGCTGGGATACAGGAGTGCAGAGGCCCAAGGGCCCCC 540
DB 1757 CCTCGGGAGTGGCAGGCTTGGCAGGAGTGCAGAGGATGCTTGGGCTTAAGGAGCACT 1816
QY 541 GGCCTCTGCTGGCCATCAGAGAGGAGTGGTCCCTTGCCCTGAGCAATGAGCCAGCCG 600
DB 1817 GGCCTCTGCTGGCCCTCAGAGAGCAATGAGGAGCTTGGCTCTGAGCAATGAGCCAGC 1876
QY 601 GCAAGCAGAGAGCAATGGCTGCGCGCTCACTGGAAGAACTTACAGAGCAAAATGCTACT 660
DB 1877 GCAATCAGAGAGTCAAGGAGTCCGCTCACTGGAAGAACTTACAGATTAATGCTACT 1936
QY 661 TTTTCTAGTGAAGAAGATTTTGTAGATGCAAGCTTTCTGTGTAAGAGCAAGTCTCA 720
DB 1937 TTTTCTAGTGAAGAAGATTTTGTAGATGCTAAGCTTTCTGTGTAAGAGCAATCTTCC 1996

```

```

QY 721 CATCTGTTTTCATTAACACTAGAGAGAACAGCAATGATAAAAAACAGATGTAGGG 780
DB 1997 CATCTGTTTTCATTAACACTAGAGAGAACAGCAATGATAAAAAACAGATGTAGGG 2056
QY 781 AGAGAGAGCCACGATGATGGCCCTCACAGACTCAGACCGTCAAAATGAATGAGTGGCTG 840
DB 2057 AGAGAGAGCCATGATGATGGCCCTCACAGACTCAGAACAGAGAGCAATGAGTGGCT 2116
QY 841 GATGGACATCTCCACACTVCAAAAAATTTGGAAGCTGAGACGCGGATTAATGGGGTCA 900
DB 2117 GATGGGTCACCTTTGATTTACAAAACTGGAAGCTGAGAACCAATTAATGGGGCAGT 2176
QY 901 GGCATGAGGCCAGAGAGACTGCTGCTGGTGTATTTATGCTGGGAGTGAACGATTTTC 960
DB 2177 GGCATGAGGCCAGAGAGACTGCTGCTGGTGTATTTATGCTGGGAGTGAACGATTTTC 2236
QY 961 CATGTGAAAGCTCAATTAATCTTATTTGGAAAAAGACGGAGACAGTATGTATCT 1020
DB 2237 CATGTGATGAATCAATTAATCTTATTTGGAAAGGAGGAGGAGCAGTATCATCTCC 2296
QY 1021 GCATTA 1026
DB 2297 ATATTA 2302
QY 1021 GCATTA 1026
DB 2297 ATATTA 2302
RESULT 9
AB052103 4330 bp mRNA linear PRI 08-MAR-2001
LOCUS AB052103
DEFINITION Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
ACCESSION AB052103
VERSION AB052103.1 GI:13365552
KEYWORDS
SOURCE
ORGANISM Homo sapiens
CDNA to mRNA.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Nakamura,K., Funkkoshi,H., Miyamoto,K., Tokunaga,F. and Nakamura,T.
TITLE Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRCL), a novel member of a
scavenger receptor family
JOURNAL Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
MEDLINE 21092718
REFERENCE 2 (bases 1 to 4330)
AUTHORS Nakamura,K. and Nakamura,T.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-2000) Kenji Nakamura, Osaka University Graduate
School of Medicine, Division of Biochemistry, Biomedical Research
Center, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail:knakamura@bich.med.osaka-u.ac.jp, Tel:81-6-6879-3783,
Fax:81-6-6879-3789)
FEATURES
source
1..4330
location/Qualifiers
1..4330
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..4330
/gene="SRCL"
3..1871
/gene="SRCL"
/codon_start=1
/product="scavenger receptor with C-type lectin type II"
/protein_id="BAB39148.1"
/db_xref="GI:13365553"
/translation="MKDPAEEBEVPSFGFRFGIHEGTQCTKCIINMALKEFISVILY
ILCALITVITVILGKVVKEKMDNVSQDGTETSKNDTLKLAQSDALVDSRSLKEDGAGKAI
SNSELSFTRSDIIDLROOLITETKTSKNDTLKLAQSDALVDSRSLKEDGAGKAI
SPILITVVKTLQAIYNGYTNLQODITVNOGINSOSVYIMNLINLNUVOQRN
LISNLQSDVDTSLAIORIKNDPQNLQOVFLQAKDIDWLKEKVSLOTLANNSALA
KANNDTLEDMSQLSFTGQMDNITTTISQANDQSLKLDLDRKDENRVAFSOLEE
RPOVETDVIINISYTAHRLRTLSMNDVTTCTDITLRTDIDLTLNNTLVNI
RDSISLRQODMRSKLDTEVANSVVEEMKLVDSKGLINKNTILQGPGRGP
KDRGSQGPPTGKNGKQKGEKPEPGEKRGITGVPGSEVDSKSGSGSPK
GSRGSPKRGPGSPGDRGPPGPKDGLPGQGPGRGLOGTGSEVPVPRGLPG
LPVPGMPKPPKPPPPPSGMBELADINEPTVASEVGCPPHMKNTFDKCTYSLE
KEILDALKFECDKSHLVFINSREQDQWIKHVTYGRSHMIGLADSEQSEMKWLDG
SPVDYKMKMAGQPDNMWGSQHGPEDCAGLIVAGWNDFQCDINNEICKEFEAVPSS
IL"

```



```

VERSION      AC024368.1 GI:7108157
KEYWORDS     HTG: HTGS_PHASE0.
SOURCE       Homo sapiens.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE    1 (bases 1 to 71044)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE        Homo sapiens chromosome 11, clone RP11-179K3
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 71044)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
              Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavskiy,L.,
              Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
              Chepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
              DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
              Fennestor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
              Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,
              Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
              Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
              Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
              Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
              McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
              Menous,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J.,
              Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivier,T.M.,
              Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C.,
              Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
              Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
              Subramanian,A., Talamas,J., Testfaye,S., Theodore,J., Tirelli,A.,
              Travers,M., Triggillo,J., Vassiliou,H., Viel,R., Vo,A., Wilson,B.,
              Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
              Zody,M.

TITLE        Direct Submission
JOURNAL      Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      All repeats were identified using RepeatMasker:
              Smit, A.F.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html

              ----- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
              ----- Project Information
              Center project name: L6145
              Center clone name: 179_K_3

* NOTE: This record contains 90 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
  
```

```

* 6184 6283: gap of 100 bp
* 6284 6984: contig of 701 bp in length
* 6985 7084: gap of 100 bp
* 7085 7768: contig of 684 bp in length
* 7769 7868: gap of 100 bp
* 7869 8550: contig of 682 bp in length
* 8551 8650: gap of 100 bp
* 8651 9352: contig of 702 bp in length
* 9353 9452: gap of 100 bp
* 9453 10140: contig of 688 bp in length
* 10141 10240: gap of 100 bp
* 10241 10922: contig of 682 bp in length
* 10923 11022: gap of 100 bp
* 11023 11709: contig of 687 bp in length
* 11710 11809: gap of 100 bp
* 11810 12511: contig of 702 bp in length
* 12512 12611: gap of 100 bp
* 12612 13304: contig of 693 bp in length
* 13305 13404: gap of 100 bp
* 13405 14095: contig of 691 bp in length
* 14096 14195: gap of 100 bp
* 14196 14881: contig of 686 bp in length
* 14882 14981: gap of 100 bp
* 14982 15674: contig of 693 bp in length
* 15675 15774: gap of 100 bp
* 15775 16460: contig of 686 bp in length
* 16461 16560: gap of 100 bp
* 16561 17250: contig of 690 bp in length
* 17251 17350: gap of 100 bp
* 17351 18040: contig of 690 bp in length
* 18041 18140: gap of 100 bp
* 18141 18836: contig of 696 bp in length
* 18837 18936: gap of 100 bp
* 18937 19621: contig of 686 bp in length
* 19622 19721: gap of 100 bp
* 19722 20425: contig of 704 bp in length
* 20426 20525: gap of 100 bp
* 20526 21224: contig of 699 bp in length
* 21225 21324: gap of 100 bp
* 21325 22017: contig of 693 bp in length
* 22018 22117: gap of 100 bp
* 22118 22815: contig of 696 bp in length
* 22816 22915: gap of 100 bp
* 22916 23609: contig of 694 bp in length
* 23610 23709: gap of 100 bp
* 23710 24402: contig of 693 bp in length
* 24403 24502: gap of 100 bp
* 24503 25191: contig of 689 bp in length
* 25192 25291: gap of 100 bp
* 25292 25981: contig of 690 bp in length
* 25982 26081: gap of 100 bp
* 26082 26774: contig of 693 bp in length
* 26775 26874: gap of 100 bp
* 26875 27563: contig of 689 bp in length
* 27564 27663: gap of 100 bp
* 27664 28355: contig of 692 bp in length
* 28356 28455: gap of 100 bp
* 28456 29145: contig of 690 bp in length
* 29146 29245: gap of 100 bp
* 29246 29943: contig of 696 bp in length
* 29944 30043: gap of 100 bp
* 30044 30745: contig of 702 bp in length
* 30746 30845: gap of 100 bp
* 30846 31535: contig of 690 bp in length
* 31536 31635: gap of 100 bp
* 31636 32329: contig of 694 bp in length
* 32330 32429: gap of 100 bp
* 32430 33114: contig of 685 bp in length
* 33115 33214: gap of 100 bp
* 33215 33898: contig of 684 bp in length
* 33899 33998: gap of 100 bp
* 33999 34694: contig of 696 bp in length
* 34695 34794: gap of 100 bp
  
```

| | | | |
|-------|--------|------------------|-----------|
| 34795 | 35486: | contlg of 692 bp | in length |
| 35487 | 35586: | gap of 100 bp | |
| 35587 | 36292: | contlg of 706 bp | in length |
| 36293 | 36392: | gap of 100 bp | |
| 36393 | 37095: | contlg of 703 bp | in length |
| 37096 | 37195: | gap of 100 bp | |
| 37196 | 37872: | contlg of 677 bp | in length |
| 37873 | 37972: | gap of 100 bp | |
| 37973 | 38663: | contlg of 691 bp | in length |
| 38664 | 38763: | gap of 100 bp | |
| 38764 | 39449: | contlg of 686 bp | in length |
| 39450 | 39549: | gap of 100 bp | |
| 39550 | 40236: | contlg of 687 bp | in length |
| 40237 | 40336: | gap of 100 bp | |
| 40337 | 41025: | contlg of 689 bp | in length |
| 41026 | 41125: | gap of 100 bp | |
| 41126 | 41818: | contlg of 693 bp | in length |
| 41819 | 41918: | gap of 100 bp | |
| 41919 | 42608: | contlg of 690 bp | in length |
| 42609 | 42708: | gap of 100 bp | |
| 42709 | 43399: | contlg of 691 bp | in length |
| 43400 | 43499: | gap of 100 bp | |
| 43500 | 44109: | contlg of 670 bp | in length |
| 44170 | 44289: | gap of 100 bp | |
| 44270 | 44966: | contlg of 697 bp | in length |
| 44967 | 45066: | gap of 100 bp | |
| 45067 | 45750: | contlg of 684 bp | in length |
| 45751 | 45850: | gap of 100 bp | |
| 45851 | 46545: | contlg of 695 bp | in length |
| 46546 | 46645: | gap of 100 bp | |
| 46646 | 47336: | contlg of 691 bp | in length |
| 47337 | 47436: | gap of 100 bp | |
| 47437 | 48123: | contlg of 687 bp | in length |
| 48124 | 48223: | gap of 100 bp | |
| 48224 | 48899: | contlg of 676 bp | in length |
| 48900 | 48999: | gap of 100 bp | |
| 49000 | 49672: | contlg of 673 bp | in length |
| 49673 | 49772: | gap of 100 bp | |
| 49773 | 50465: | contlg of 693 bp | in length |
| 50466 | 50565: | gap of 100 bp | |
| 50566 | 51254: | contlg of 689 bp | in length |
| 51255 | 51354: | gap of 100 bp | |
| 51355 | 52062: | contlg of 708 bp | in length |
| 52063 | 52162: | gap of 100 bp | |
| 52163 | 52847: | contlg of 685 bp | in length |
| 52848 | 52947: | gap of 100 bp | |
| 52948 | 53646: | contlg of 699 bp | in length |
| 53647 | 53746: | gap of 100 bp | |
| 53747 | 54409: | contlg of 663 bp | in length |
| 54410 | 54509: | gap of 100 bp | |

| Query Match | 47.2% | Score 484.6 | DB 2 | Length 71044 |
|-----------------------|---|--------------------|----------|--------------|
| Best Local Similarity | 98.2% | Pred. No. 3.5e+101 | | |
| Matches 490 | Conservative 0 | Mismatches 9 | Indels 0 | Gaps 0 |
| QY 119 | TACTACAAGGTCACACGGGGCCACAGGGGTCTCAAGAGGTGACAGAGATCCACGGAGCCCC | | | 178 |
| Db 43599 | TAGCAAAAAGTCTCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCACGGAGCCCC | | | 43658 |
| QY 179 | CTGGCCCCAAGTGTGGCACAACAAGGACAGAAAGAGAGAAAGGGGGAGCTGTGACCACTGTGGCC | | | 238 |
| Db 43659 | CTGGCCCCAAGTGTGGCACAACAAGGACAGAAAGAGAGAAAGGGGGAGCTGTGACCACTGTGGCC | | | 43718 |
| QY 239 | CTGGGGGTGAGAGAGAGGCCCAATTGGACACAGCTGTGCCCCCGGAGAGGGGTGGCGGGCAAG | | | 298 |
| Db 43719 | CTGGGGGTGAGAGAGGCCCAAGGGGACCAAGCTGTGCCCCCGGAGAGGGGTGGCGGGCAAG | | | 43778 |
| QY 299 | GATCTAAAGGCTCCACGAGGCCCAAGAGCTCCCGTGTTCCTCTGGGAAGCCGGGCGCTTC | | | 358 |
| Db 43779 | GATCTAAAGGCTCCACGAGGCCCAAGAGCTCCCGTGTTCCTCTGGGAAGCCGGGCGCTTC | | | 43838 |
| QY 359 | AGGGCCCCAGTGTGGGACCCAGGCCCCCGGGGCCACAGCGCAAGAGGACTTCCCGGGCC | | | 418 |

| | | | |
|----|-------|--|-------|
| Db | 43839 | AGGCTCTCAGTGTGGGGACCCAGGCCCCCGGGGCCACACAGGCAAAAGAGGACTCCCGGCC | 43898 |
| QY | 419 | CTCAGGGCCCTCTCTGCTTCCAGGACTTTCAGGGCACCGCTTTGGGAGCCTGGGGTCTCTG | 478 |
| Db | 43899 | CTCAGGGCCCTCTCTGCTTCCAGGACTTTCAGGGCACCGCTTTGGGAGCCTGGGGTCTCTG | 43958 |
| QY | 479 | GACCTTGGGGAGCTGCAGAGCTTGCTGGGGGATACAGGACATGSCCAGGCCCCCAAGGGGCCCC | 538 |
| Db | 43359 | GACCTTGGGGAGCTGCAGAGCTTGCTGGGGGATACAGGACATGSCCAGGCCCCCAAGGGGCCCC | 44018 |
| QY | 539 | CGGGCCCTCTCTGAGCCATCAGAGAGCGGTGTGTCCTTGCCCTTGCCAGATGAGCCAAACC | 598 |
| Db | 44019 | CGGGCCCTCTCTGAGCCATCAGAGAGCGGTGTGTCCTTGCCCTTGCCAGATGAGCCAAACC | 44078 |
| QY | 599 | CGGCACCGGAGGACAATGG | 617 |
| Db | 44079 | CACGACCGGAGGACAATGG | 44097 |

| RESULT 13 | AP001022 | LOCUS | DEFINITION | ACCESSION | VERSION | KEYWORDS | SOURCE | ORGANISM |
|-----------|------------|-------------------------------------|---|-----------|-----------------|----------|--------|----------|
| | AP001022 | 187635 bp | DNA | Linear | HTG 30-MAY-2000 | | | |
| | | | Homo sapiens chromosome 18 clone RP11-815L4 map 18p11.3, WORKING DRAFT SEQUENCE, 33 unordered pieces. | AP001022 | | | | |
| | AP001022.2 | GI:8117692 | | | | | | |
| | | HTG: HTGS_PHASE1 | HTGS DRAFT. | | | | | |
| | | Homo sapiens DNA, clone:RP11-815L4. | | | | | | |
| | | Homo sapiens | | | | | | |

REFERENCE 1 (bases 1 to 187635)
AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Homo sapiens 187,635 genomic DNA of 18p11.3
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 187635)
AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-2000) Masahira Hattori, The Institute of Physical

COMMENT On May 31, 2000 this sequence version replaced gi:6997772

Center: RIKEN Genomic Sciences Center (GSC)

Center code: RIKEN

Web site: <http://hgp.gsc.riken.go.jp/>
Contact: hattori@gsc.riken.go.jp

```
----- Project Information
Center project name: HumDraft18
```

Center clone name: RP11-815L4

Summary Statistics

Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 167192 bases at least Q40

Consensus quality: 177020 bases at least Q30

Consensus quality: 1821.05 bases at

Insert size: 184435; sum-of-contigs
Quality coverage: 4.35x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 21332 contig of 21332 bp in length

```

21433      42072 contig of 20640 bp in length
42173      54336 contig of 12164 bp in length
54437      66766 contig of 12330 bp in length
66867      77849 contig of 10983 bp in length
77950      87798 contig of 9849 bp in length
87899      99510 contig of 11612 bp in length
99611      107273 contig of 7663 bp in length
107374      113189 contig of 5816 bp in length
113290      120559 contig of 7270 bp in length
120660      126468 contig of 5809 bp in length
126569      131619 contig of 5051 bp in length
131720      136810 contig of 5091 bp in length
136911      139925 contig of 3015 bp in length
140026      144232 contig of 4207 bp in length
144333      148902 contig of 4570 bp in length
149003      152918 contig of 3916 bp in length
153019      156400 contig of 2434 bp in length
156500      158934 contig of 100 bp in length
158934      159033 contig of 100 bp in length
159034      161884 contig of 2851 bp in length
161885      161984 contig of 100 bp in length
161985      164301 contig of 2317 bp in length
164402      166300 contig of 1899 bp in length
166401      168580 contig of 2180 bp in length
168681      171270 contig of 2590 bp in length
171371      174092 contig of 2722 bp in length
174193      176095 contig of 1903 bp in length
176196      177890 contig of 1695 bp in length
177991      179786 contig of 1796 bp in length
179887      181478 contig of 1592 bp in length
181579      182896 contig of 1317 bp in length
182996      184841 contig of 1846 bp in length
184942      186345 contig of 1404 bp in length
186446      187635 contig of 1190 bp in length

```

Sequence updated (26-May-2000)

* NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1      21332 contig of 21332 bp in length
21333      21432 gap of 100 bp
21433      42072 contig of 20640 bp in length
42073      42172 gap of 100 bp
42173      54336 contig of 12164 bp in length
54337      54436 gap of 100 bp
54437      66766 contig of 12330 bp in length
66767      66866 gap of 100 bp
66867      77849 contig of 10983 bp in length
77850      77949 gap of 100 bp
77950      87798 contig of 9849 bp in length
87799      87898 gap of 100 bp
87899      99510 contig of 11612 bp in length
99511      99610 gap of 100 bp
99611      107273 contig of 7663 bp in length
107274      107373 gap of 100 bp
107374      113189 contig of 5816 bp in length
113190      113289 gap of 100 bp
113290      120559 contig of 7270 bp in length
120560      120659 gap of 100 bp
120660      126468 contig of 5809 bp in length
126469      126568 gap of 100 bp
126569      131619 contig of 5051 bp in length
131620      131719 gap of 100 bp
131720      136810 contig of 5091 bp in length
136811      136910 gap of 100 bp
136911      139925 contig of 3015 bp in length
139926      140025 gap of 100 bp
140026      144232 contig of 4207 bp in length
144233      144332 gap of 100 bp
144333      148902 contig of 4570 bp in length
148903      149002 gap of 100 bp

```

```

*      149003      152918: contig of 3916 bp in length
*      152919      153018: gap of 100 bp
*      153019      156399: contig of 3381 bp in length
*      156400      156499: gap of 100 bp
*      156500      158933: contig of 2434 bp in length
*      158934      159033: gap of 100 bp
*      159034      161884: contig of 2851 bp in length
*      161885      161984: gap of 100 bp
*      161985      164301: contig of 2317 bp in length
*      164302      164401: gap of 100 bp
*      164402      166300: contig of 1899 bp in length
*      166301      166400: gap of 100 bp
*      166401      168580: contig of 2180 bp in length
*      168581      168680: gap of 100 bp
*      168681      171270: contig of 2590 bp in length
*      171271      171370: gap of 100 bp
*      171371      174092: contig of 2722 bp in length
*      174093      174192: gap of 100 bp
*      174193      176095: contig of 1903 bp in length
*      176096      176195: gap of 100 bp
*      176196      177890: contig of 1695 bp in length
*      177891      177990: gap of 100 bp
*      177991      179786: contig of 1796 bp in length
*      179787      179886: gap of 100 bp
*      179887      181478: contig of 1592 bp in length
*      181479      181578: gap of 100 bp
*      181579      182895: contig of 1317 bp in length
*      182896      182995: gap of 100 bp
*      182996      184841: contig of 1846 bp in length
*      184842      184941: gap of 100 bp
*      184942      186345: contig of 1404 bp in length
*      186346      186445: gap of 100 bp
*      186446      187635: contig of 1190 bp in length.

```

FEATURES

source

```

misc_feature      1..21332
                    /note="assembly-fragment"
                    /db_xref="taxon:9606"
                    /map="18p11.3"
                    /clone="Rp11-815L4"
                    1..21332
misc_feature      21433..42072
                    /note="assembly-fragment"
                    /note="assembly-fragment"
                    42173..54336
                    /note="assembly-fragment"
                    /note="assembly-fragment"
                    54437..66766
                    /note="assembly-fragment"
                    /note="assembly-fragment"
                    66867..77849
                    /note="assembly-fragment"
                    77950..87798
                    /note="assembly-fragment clone-end:T7 vector-side:right"
                    87899..99510
                    /note="assembly-fragment"
                    /note="assembly-fragment"
                    99611..107273
                    /note="assembly-fragment"
                    107374..113189
                    /note="assembly-fragment clone-end:SP6 vector-side:left"
                    113290..120559
                    /note="assembly-fragment"
                    120660..126468
                    /note="assembly-fragment"
                    126569..131619
                    /note="assembly-fragment"
                    131720..136810
                    /note="assembly-fragment"
                    136911..139925
                    /note="assembly-fragment"
                    140026..144232
                    /note="assembly-fragment"
                    144333..148902
                    /note="assembly-fragment"
                    149003..152918
                    misc_feature

```



```
Center clone name: RP11-839023
----- Summary Statistics
Sequencing vector: PCR products: 100% of reads
Chemistry: Dye-terminator EF-amerisam: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 171068 bases at least Q40
Consensus quality: 178865 bases at least Q30
Consensus quality: 182445 bases at least Q20
Insert size: 185155; sum-of-ctrls
Quality coverage: 4.41x in Q20 bases; sum-of-ctrls
-----
NOTE: This is a 'working draft' sequence. It currently consists of
32 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1      16096 contig of 16096 bp in length
16197 33789 contig of 17593 bp in length
33890 51010 contig of 17121 bp in length
51111 63558 contig of 12448 bp in length
63559 75056 contig of 11398 bp in length
75157 84737 contig of 9581 bp in length
84838 90909 contig of 6072 bp in length
90910 98615 contig of 7606 bp in length
98716 105978 contig of 7263 bp in length
106079 113049 contig of 6971 bp in length
113150 119201 contig of 6052 bp in length
119302 126582 contig of 7281 bp in length
126683 132070 contig of 5388 bp in length
132171 137138 contig of 4968 bp in length
137139 137238 contig of 100 bp in length
137239 143055 contig of 5811 bp in length
143056 143155 contig of 100 bp in length
143156 148639 contig of 5484 bp in length
148640 148739 contig of 100 bp in length
148740 153086 contig of 4347 bp in length
153087 153186 contig of 100 bp in length
153187 158921 contig of 5735 bp in length
158922 159021 contig of 100 bp in length
159022 163174 contig of 4153 bp in length
163175 163274 contig of 100 bp in length
163275 167066 contig of 3792 bp in length
167067 167166 contig of 100 bp in length
167167 170120 contig of 2954 bp in length
170121 170220 contig of 100 bp in length
170221 172430 contig of 2210 bp in length
172431 172530 contig of 100 bp in length
172531 175598 contig of 3068 bp in length
175599 175699 contig of 1420 bp in length
175699 177118 contig of 100 bp in length
177119 177218 contig of 1313 bp in length
177219 178531 contig of 100 bp in length
178532 178631 contig of 100 bp in length
178632 179857 contig of 1226 bp in length
179858 179957 contig of 100 bp in length
179958 181064 contig of 1107 bp in length
181065 181164 contig of 100 bp in length
181165 182735 contig of 1571 bp in length
182736 182835 contig of 100 bp in length
182836 183923 contig of 1088 bp in length
183924 184023 contig of 100 bp in length
184024 185810 contig of 1787 bp in length
185811 185910 contig of 100 bp in length
185911 187145 contig of 1235 bp in length
187146 187245 contig of 100 bp in length
187246 188255 contig of 1010 bp in length
Sequence updated (06-Jan-2000)
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1      16096 contig of 16096 bp in length
16097 16196 contig of 100 bp in length
16197 33789 contig of 17593 bp in length
33890 51010 contig of 17121 bp in length
51111 63558 contig of 12448 bp in length
63559 75056 contig of 11398 bp in length
75157 84737 contig of 9581 bp in length
84838 90909 contig of 6072 bp in length
90910 98615 contig of 7606 bp in length
```

```
90910 91009: gap of 100 bp
91010 98615: contig of 7606 bp in length
98616 98715: gap of 100 bp
98716 105978: contig of 7263 bp in length
105979 106078: gap of 100 bp
106079 113049: contig of 6971 bp in length
113050 113149: gap of 100 bp
113150 119201: contig of 6052 bp in length
119202 119301: gap of 100 bp
119302 126582: contig of 7281 bp in length
126583 126682: gap of 100 bp
126683 132070: contig of 5388 bp in length
132071 132170: gap of 100 bp
132171 137138: contig of 4968 bp in length
137139 137238: gap of 100 bp
137239 143055: contig of 5811 bp in length
143056 143155: gap of 100 bp
143156 148639: contig of 5484 bp in length
148640 148739: gap of 100 bp
148740 153086: contig of 4347 bp in length
153087 153186: gap of 100 bp
153187 158921: contig of 5735 bp in length
158922 159021: gap of 100 bp
159022 163174: contig of 4153 bp in length
163175 163274: gap of 100 bp
163275 167066: contig of 3792 bp in length
167067 167166: gap of 100 bp
167167 170120: contig of 2954 bp in length
170121 170220: gap of 100 bp
170221 172430: contig of 2210 bp in length
172431 172530: gap of 100 bp
172531 175598: contig of 3068 bp in length
175599 175699: gap of 100 bp
175699 177118: contig of 1420 bp in length
177119 177218: gap of 100 bp
177219 178531: contig of 1313 bp in length
178532 178631: gap of 100 bp
178632 179857: contig of 1226 bp in length
179858 179957: gap of 100 bp
179958 181064: contig of 1107 bp in length
181065 181164: gap of 100 bp
181165 182735: contig of 1571 bp in length
182736 182835: gap of 100 bp
182836 183923: contig of 1088 bp in length
183924 184023: gap of 100 bp
184024 185810: contig of 1787 bp in length
185811 185910: gap of 100 bp
185911 187145: contig of 1235 bp in length
187146 187245: gap of 100 bp
187246 188255: contig of 1010 bp in length.
FEATURES
source
1..188255
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-839023"
1..16096
/feature="assembly_fragment"
16197..33789
/feature="assembly_fragment"
33890..51010
/feature="assembly_fragment"
51111..63558
/feature="assembly_fragment"
63559..75056
/feature="assembly_fragment"
75157..84737
/feature="assembly_fragment"
84838..90909
/feature="assembly_fragment"
91010..98615
/feature="assembly_fragment"
```

```
misc_feature      98716..105978
                  /note="assembly_fragment"
misc_feature      106079..113049
                  /note="assembly_fragment"
misc_feature      113150..119201
                  /note="assembly_fragment"
misc_feature      119302..126582
                  /note="assembly_fragment"
misc_feature      126683..132070
                  /note="assembly_fragment"
misc_feature      132171..137138
                  /note="assembly_fragment"
misc_feature      137239..143055
                  /note="assembly_fragment"
misc_feature      143156..148639
                  /note="assembly_fragment"
misc_feature      148740..153086
                  /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature      153187..158921
                  /note="assembly_fragment"
misc_feature      159022..163174

Query Match      42.6%; Score 437.4; DB 2; Length 188255;
Best Local Similarity 81.3%; Pred.No.2.9e-90;
Matches 438; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 126 AGGTCCACCGGGGCCAGGGGTCCAGAGGTGACAGAGATCCAGGACCCCTGGGCC 185
      |||||||
Db 184461 AGGTCCACCGGGGCCAGGGGTCCAGAGGTGACAGAGATCCAGGACCCCTGGGCC 184402

QY 186 AACTGGCAACAAGGACAGAAAGAGAGAAGGGGAGCCTGGACACCTGGCCCTCGGG 245
      |||||||
Db 184401 AACTGGCAACAAGGACAGAAAGAGAGAAGGGGAGCCTGGACACCTGGCCCTCGGG 184342

QY 246 TGAGAGAGGCCCAATTGGACACAGCTGTCCCGGAGAGCTGGCGGCAAGATCTAA 305
      |||||||
Db 184341 TGAGAGAGGCCCAATTGGACACAGCTGTCCCGGAGAGCTGGCGGCAAGATCTAA 184282

QY 306 AGGCTCCAGGGGCCCAAGGCTCCCGTGGTCCCTGGGAAAGCCCGGCCCTCAGGGCC 365
      |||||||
Db 184281 AGGCTCCAGGGGCCCAAGGCTCCCGTGGTCCCTGGGAAAGCCCGGCCCTCAGGGCTC 184222

QY 366 CAGTGGGACCCAGGCCCCCGGCCACACAGGCAAAAGAGGACTCCCGGCCCTCAGGG 425
      |||||||
Db 184221 CAGTGGGACCCAGGCCCCCGGCCACACAGGCAAAAGAGGACTCCCGGCCCTCAGGG 184162

QY 426 CCTCTCGGCTTCCAGGACTTCAGGGACCGTTGGGAGCCTGGGGTGCCTGGACTCG 485
      |||||||
Db 184161 CCTCTCGGCTTCCAGGACTTCAGGGACCGTTGGGAGCCTGGGGTGCCTGGACTCG 184102

QY 486 GGGACTGCCAGGCTGCTGGGGTACAGGATGCCAGGCCCAAGGGCCCCCGGGCC 545
      |||||||
Db 184101 GGGACTGCCAGGCTGCTGGGGTACAGGATGCCAGGCCCAAGGGCCCCCGGGCC 184042

QY 546 TCCTGGCCATCAGAGCGGTGCTGCCCTGGCCCTGCAGAATGAGCAACCCCGGCACC 605
      |||||||
Db 184041 TCCTGGCCATCAGAGCGGTGCTGCCCTGGCCCTGCAGAATGAGCAACCCCGGCACC 183982

QY 606 GGAAGACAATGGCTGCCCTCAGTGAAGACTTCACAGACAATGCTACTATTTT 664
      |||||||
Db 183981 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNT 183923
```

Search completed: March 21, 2003, 06:43:48
Job time : 3104.5 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 02:00:05 ; Search time 36.1652 Seconds
(Without alignments)
8700.356 Million cell updates/sec

Title: US-09-763-712a-1_COPY-670_1695
Perfect score: 1026
Sequence: 1 atgcacacagattgatgag.....cagctactcatctgcatctta 1026

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_NA:*
2: /cgn2_6/ptodata/1/lna/5A_COMB.seq:*
3: /cgn2_6/ptodata/1/lna/5B_COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6A_COMB.seq:*
5: /cgn2_6/ptodata/1/lna/6B_COMB.seq:*
6: /cgn2_6/ptodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Query length | DB ID | Description |
|------------|-------|-------------|--------------|-------|--------------------------------------|
| 1 | 121.6 | 11.9 | 810 | 1 | US-08-642-255-60 Sequence 60, Appl |
| 2 | 118 | 11.5 | 756 | 1 | US-08-642-255-50 Sequence 50, Appl |
| 3 | 113.6 | 11.1 | 3181 | 1 | US-08-655-086-1 Sequence 1, Appl |
| 4 | 106 | 10.3 | 432 | 1 | US-08-642-255-48 Sequence 48, Appl |
| 5 | 105 | 10.3 | 1608 | 4 | US-09-029-348-19 Sequence 19, Appl |
| 6 | 100 | 9.7 | 5102 | 1 | US-08-484-168-1 Sequence 1, Appl |
| 7 | 99.6 | 9.7 | 1416 | 1 | US-07-621-0916-1 Sequence 1, Appl |
| 8 | 99.6 | 9.7 | 1416 | 2 | US-08-399-889-1 Sequence 1, Appl |
| 9 | 99.6 | 9.7 | 1416 | 3 | US-09-167-364-1 Sequence 1, Appl |
| 10 | 99.6 | 9.7 | 1416 | 4 | US-09-439-897-1 Sequence 1, Appl |
| 11 | 99.4 | 9.7 | 1868 | 1 | US-08-392-3678-1 Sequence 1, Appl |
| 12 | 99.4 | 9.7 | 1868 | 3 | US-08-893-467A-1 Sequence 1, Appl |
| 13 | 99.4 | 9.7 | 2343 | 1 | US-08-535-669-11 Sequence 11, Appl |
| 14 | 99.4 | 9.7 | 2543 | 3 | US-09-073-663-11 Sequence 11, Appl |
| 15 | 97.6 | 9.5 | 4359 | 4 | US-09-484-9708-4 Sequence 4, Appl |
| 16 | 96.2 | 9.4 | 1572 | 4 | US-09-297-269-39 Sequence 5, Appl |
| 17 | 95.2 | 9.3 | 1560 | 4 | US-08-794-795-5 Sequence 5, Appl |
| 18 | 95.2 | 9.3 | 1560 | 4 | US-08-249-200-5 Sequence 5, Appl |
| 19 | 95.2 | 9.3 | 1703 | 4 | US-08-794-795-1 Sequence 1, Appl |
| 20 | 95.2 | 9.3 | 1703 | 4 | US-09-249-200-1 Sequence 1, Appl |
| 21 | 89 | 8.7 | 1881 | 4 | US-09-029-348-20 Sequence 20, Appl |
| 22 | 84.4 | 8.2 | 506 | 1 | US-08-253-155A-61 Sequence 61, Appl |
| 23 | 77 | 7.5 | 1560 | 4 | US-09-453-702B-264 Sequence 264, App |
| 24 | 77 | 7.5 | 61663 | 4 | US-09-453-702B-62 Sequence 62, Appl |
| 25 | 76.8 | 7.5 | 5552 | 3 | US-08-155-888-1 Sequence 1, Appl |
| 26 | 74.6 | 7.3 | 38584 | 4 | US-09-453-702B-50 Sequence 50, Appl |
| 27 | 73 | 7.1 | 503 | 4 | US-09-297-269-40 Sequence 40, Appl |

| | | | | | | | |
|----|------|------|------|-------|--------------------|--------------------|--------------------|
| 28 | 72.4 | 7.1 | 4031 | 1 | US-08-159-784-1 | Sequence 1, Appl | |
| c | 29 | 72.2 | 7.0 | 9837 | 4 | US-09-453-702B-66 | Sequence 66, Appl |
| 30 | 71.4 | 7.0 | 392 | 4 | US-09-404-879A-372 | Sequence 372, Appl | |
| 31 | 70 | 6.8 | 513 | 4 | US-09-134-001C-647 | Sequence 647, Appl | |
| c | 32 | 70 | 6.8 | 585 | 4 | US-09-134-001C-726 | Sequence 726, Appl |
| c | 33 | 70 | 6.8 | 810 | 4 | US-09-134-001C-624 | Sequence 624, Appl |
| c | 34 | 70 | 6.8 | 3552 | 4 | US-09-134-001C-693 | Sequence 693, Appl |
| c | 35 | 69 | 6.7 | 1347 | 4 | US-09-140-804-1 | Sequence 1, Appl |
| c | 36 | 68.6 | 6.7 | 45175 | 4 | US-09-453-702B-116 | Sequence 116, Appl |
| c | 37 | 67.8 | 6.6 | 3404 | 4 | US-09-453-702B-94 | Sequence 94, Appl |
| c | 38 | 67.8 | 6.6 | 48908 | 4 | US-09-453-702B-137 | Sequence 137, Appl |
| c | 39 | 67.2 | 6.5 | 729 | 4 | US-09-140-804-10 | Sequence 10, Appl |
| 40 | 67 | 6.5 | 226 | 1 | US-08-175-155-63 | Sequence 63, Appl | |
| 41 | 67 | 6.5 | 226 | 1 | US-08-477-509B-98 | Sequence 98, Appl | |
| 42 | 67 | 6.5 | 226 | 1 | US-08-642-255-96 | Sequence 96, Appl | |
| 43 | 67 | 6.5 | 226 | 1 | US-08-397-633A-16 | Sequence 16, Appl | |
| 44 | 67 | 6.5 | 226 | 2 | US-08-707-237A-70 | Sequence 70, Appl | |
| 45 | 67 | 6.5 | 226 | 3 | US-08-482-085B-98 | Sequence 98, Appl | |

ALIGNMENTS

```
RESULT 1
US-08-642-255-60
: Sequence 60, Application US/08642255
: Patent No. 5773249
:
: GENERAL INFORMATION:
: APPLICANT: CAPPELLO, Joseph
: APPLICANT: FERRARI, Franco A.
: TITLE OF INVENTION: High Molecular Weight Collagen-Like
: TITLE OF INVENTION: Protein Polymers
: NUMBER OF SEQUENCES: 135
: CORRESPONDENCE ADDRESS:
: ADDRESSSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/642,255
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: ROWLAND, Bertram I.
: REGISTRATION NUMBER: 20,015
: REFERENCE/DOCKET NUMBER: A55556-3/BIR
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 494-8700
: TELEFAX: (415) 494-8771
: TELEX: 910 277299 FRT UR
: INFORMATION FOR SEQ ID NO: 60:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 810 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "synthetic"
:
: US-08-642-255-60
:
: Query Match 11.9% Score 121.6; DB 1; Length 810;
: Best Local Similarity 53.9% Pred. No. 1.9e-23;
: Matches 250; Conservative 0; Mismatches 214; Indels 0; Gaps 0;
:
: 11 11 111 1111 11111 111111 111111 11 11111 11 11111 11
```



```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 3181 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-655-086-1

Query Match      11.1%; Score 113.6; DB 1; Length 3181;
Best Local Similarity 52.5%; Pred. No. 5e-21;
Matches 248; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY  97 CAGCTCATCAGAATTTTACAACTACTACAAGTTCACCGGGCCCCAGGGGTCCAAAGAGGT 156
    || || || || || || || || || || || || || || || || || || || || ||
Db  104 CTGTGCCCAAGGCTTCCCAAGGCTTCCAAGGCTCCCAAGGCTCCCGTGTGAGCCTGGCAGCCTGGA 163
    || || || || || || || || || || || || || || || || || || || || ||
QY  157 GACAGAGATCCCAAGGAGCCCTCGGGCCAACTGCGACAAAGGACACAAAGGAGAGAAG 216
    || || || || || || || || || || || || || || || || || || || || ||
Db  164 GCTTCAGTCCCATGGGTCTCCCGAGGTCCCCAGGTCCCCCTGGAAAAGAAATGGAGATGAT 223
    || || || || || || || || || || || || || || || || || || || || ||
QY  217 GGGGAGGCTGGACCACTGGCCCTCGCGGCTCAGAGAGGCCCAATTTGGACCAAGCTGGTGCC 276
    || || || || || || || || || || || || || || || || || || || || ||
Db  224 GGGGAGCTGGAARACCTGGTCGTCCTGGTGAGCGTGGGCCCTCCTGGGCCCTCAGGGTGCT 283
    || || || || || || || || || || || || || || || || || || || || ||
QY  277 CCCGGAGAGCGTGGCGGCAAGAGTCTAAAGGCTCCCAAGGCCCCCAAGGCTCCCGTGCT 336
    || || || || || || || || || || || || || || || || || || || || ||
Db  284 CGAGGATTGCCGGAACAGCTGGCTCCCTGGAATGAAGGCACACAGAGGTTTCAGTGTGT 343
    || || || || || || || || || || || || || || || || || || || || ||
QY  337 TCCCTGGGAGCCCGGCCCTCAGGGCCCCAGTGGGAGCCACAGGCCCCCCGGGGCCACCA 396
    || || || || || || || || || || || || || || || || || || || || ||
Db  344 TTGATGGTGCCAAGGAGATGCTGGTGCTCTGCTGCTCTTAAGGGTGAAGCCTGGCAGCCCT 403
    || || || || || || || || || || || || || || || || || || || || ||
QY  397 GGCAAGAGGAGACTCCCGGGCCCTCAGGGCCCTCTGSGCTTCCAGGGACTTCAGGGCACC 456
    || || || || || || || || || || || || || || || || || || || || ||
Db  404 GGTGAATGAGAGCTCTGTCAGATGGGCCCCCGTGGCCCTGCTGCTGAGAGAGTGCTGC 463
    || || || || || || || || || || || || || || || || || || || || ||
QY  457 GTTGGGAGCCTGGGTGCTCTGGACCTCGGGGACTGCCAGGCTTTCCTGGGGTACCAAGGC 516
    || || || || || || || || || || || || || || || || || || || || ||
Db  464 CTGGAGCCCTGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 523
    || || || || || || || || || || || || || || || || || || || || ||
QY  517 ATGCCAGGCCCAAGGGCCCCCGGGCCCTCTGCGCCCATCAGAGCGGTGG 568
    || || || || || || || || || || || || || || || || || || || || ||
Db  524 CCCCTGGTCCCAAGGGCCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 575
    || || || || || || || || || || || || || || || || || || || || ||

RESULT 4
US-08-642-255-48
; Sequence 48, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.

```


TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 972-1400
TELEFAX: (212) 370-1622
TELEX: 236268
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1416 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: No
ANTI-SENSE: No
ORIGINAL SOURCE:
ORGANISM: Calf
STRAIN: Unknown
INDIVIDUAL ISOLATE: Unknown
DEVELOPMENTAL STAGE: Unknown
CELL TYPE: Whole kidney
IMMEDIATE SOURCE:
LIBRARY: Bovine lens cDNA
CLONE: KMC15
POSITION IN GENOME: No. 5424408 known
CHROMOSOME/SEGMENT: No. 5424408 known
PUBLICATION INFORMATION: No. 5424408e
US-07-621-091G-1

Query Match 9.7%; Score 99.6; DB 1; Length 1416;
Best Local Similarity 51.6%; Pred. No. 2e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| Qy | 127 | GGTCCACCGGGCCCCAGGGGTCACAGAGGTGACAGAGATCCACAGGACCCCTGGGCCA | 186 |
| Db | 283 | GGAGCACCAGGTGTCCCGGTGACGCGGCGCAGGCGCAGAGGTATCTTGATTTATGATTT | 342 |
| Qy | 187 | ACTGCAACAGGACAGAAAGAGAGAGGGGGAGCCTGGACACCTGGCCCTCGGGT | 246 |
| Db | 343 | CCAGCATGAAGGGAAGGGTAATTCAGGATTTCCAGGACCACTGGACCTCCAGGG | 402 |
| Qy | 247 | GAGAGAGCCCAATTTGGACAGCTGGTCCCGCGAGAGCGTGGCGGCAAGGATCTAAA | 306 |
| Db | 403 | CAAGTGGACCAAAAGGACCACTGGAGTACGTGGAGAGCCTGGGCACAGTGAAGATCATC | 462 |
| Qy | 307 | GGTCCACCGGGCCCCAGGGTCCCGGTGTCCTCCCTGGGAAGCGCGCCTCAGGGCCCC | 366 |
| Db | 463 | TCCCTTCCAGGAAGCCAGGCCACCTGTTTCAGTGTGAGAACAGGATGCAAGAGAA | 522 |
| Qy | 367 | AGTGGGGACCCAGGCCCCCGGGTCCCGTGGTTCCTCCCTGGGAAGCGCGCCTCAGGGC | 426 |
| Db | 523 | CCCGGGCCCCAGGACCAACAGGAGATCCAGGACCTGTGGGCCAAAAGGTAACACGGG | 582 |
| Qy | 427 | CCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCTGGGGTGCCTGGACCTCGG | 486 |
| Db | 583 | GAGGATGTCCACAGGAACCTCTGGACCAACTGGAGAAAAAGGCAACAAGGTTGTAAA | 642 |
| Qy | 487 | GGACTGCCAGGCTTGCCTGGGTACAGGCATGCCAGGCCCAAGGCCGCCCGGCCCT | 546 |
| Db | 643 | GGAGAGCAAGGACCACTGGATCCGATGCCCTGCCAGGCTTGAAGGGGAAACCTTGAGAC | 702 |
| Qy | 547 | CCTGGCCCATCAGGACCGGTGG | 568 |
| Db | 703 | ACTGGACCACCTGCAGCAGGGG | 724 |

RESULT 8
US-08-399-889-1
Sequence 1, Application US/08399889B
Patent No. 5973120
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

FILE REFERENCE: 951263A
CURRENT APPLICATION NUMBER: US/08/399,889B
CURRENT FILING DATE: 1995-03-07
EARLIER APPLICATION NUMBER: 07/621091
EARLIER FILING DATE: 1990-11-30
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1416
TYPE: DNA
ORGANISM: Calf
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1416)
US-08-399-889-1

Query Match 9.7%; Score 99.6; DB 2; Length 1416;
Best Local Similarity 51.6%; Pred. No. 2e-17;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| Qy | 127 | GGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGATCCACAGGACCCCTGGGCCA | 186 |
| Db | 283 | GGAGCACCAGGTGTCCCGGTGACGCGGCGCAGGCGCAGAGGTATCTTGATTTATGATTT | 342 |
| Qy | 187 | ACTGCAACAGGACAGAAAGAGAGAGGGGGAGCCTGGACACCTGGCCCTCGGGT | 246 |
| Db | 343 | CCAGCATGAAGGGAAGGGTAATTCAGGATTTCCAGGACCACTGGACCTCCAGGG | 402 |
| Qy | 247 | GAGAGAGCCCAATTTGGACAGCTGGTCCCGCGAGAGCGTGGCGGCAAGGATCTAAA | 306 |
| Db | 403 | CAAGTGGACCAAAAGGACCACTGGAGTACGTGGAGAGCCTGGGCACAGTGAAGATCATC | 462 |
| Qy | 307 | GGTCCACCGGGCCCCAGGGTCCCGGTGTCCTCCCTGGGAAGCGCGCCTCAGGGCCCC | 366 |
| Db | 463 | TCCCTTCCAGGAAGCCAGGCCACCTGTTTCAGTGTGAGAACAGGATGCAAGAGAA | 522 |
| Qy | 367 | AGTGGGGACCCAGGCCCCCGGGTCCCGTGGTTCCTCCCTGGGAAGCGCGCCTCAGGGC | 426 |
| Db | 523 | CCCGGGCCCCAGGACCAACAGGAGATCCAGGACCTGTGGGCCAAAAGGTAACACGGG | 582 |
| Qy | 427 | CCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCTGGGGTGCCTGGACCTCGG | 486 |
| Db | 583 | GAGGATGTCCACAGGAACCTCTGGACCAACTGGAGAAAAAGGCAACAAGGTTGTAAA | 642 |
| Qy | 487 | GGACTGCCAGGCTTGCCTGGGTACAGGCATGCCAGGCCCAAGGCCGCCCGGCCCT | 546 |
| Db | 643 | GGAGAGCAAGGACCACTGGATCCGATGCCCTGCCAGGCTTGAAGGGGAAACCTTGAGAC | 702 |
| Qy | 547 | CCTGGCCCATCAGGACCGGTGG | 568 |
| Db | 703 | ACTGGACCACCTGCAGCAGGGG | 724 |

RESULT 9
US-09-167-364-1
Sequence 1, Application US/09167364
Patent No. 6007980
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263B
CURRENT APPLICATION NUMBER: US/09/167,364
CURRENT FILING DATE: 1998-10-07
EARLIER APPLICATION NUMBER: 08/399889
EARLIER FILING DATE: 1995-03-07
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1416
TYPE: DNA
ORGANISM: Calf

29

```
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Nucleotide-genomic DNA
; HYPOTHETICAL: No. 5691197 relevant
; ANTI-SENSE: No. 5691197 relevant
US-08-392-367B-1

Query Match          9.7%; Score 99.4; DB 1; Length 1868;
Best Local Similarity 51.5%; Pred. No. 2.6e-17;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 124 CAAGGTCCACCGGGCCCCCAGAGGTCCAAAGGTGACAGAGATCCACAGGAGCCCTCGC 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 CCAGGTGTCCAAGGACCCAGGGCCCCCAGGACCAAGGAGGAGGAGGCTCCAGGGA 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 184 CCAACTGGCAACAAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 CTTAGGGTGCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 244 GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATCT 303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 GGCAGCAAGAGTGACATAGGTCTCACTGGCCCCCAAGGGGAGACATGGCACCAAGGAGAC 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 304 AAAGGCTCCAGAGGCCCCCAAGAGTCCCGTGGTCCCTGGGAGAGCCCGCCCTCAGGGC 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 AAAGGGACCTAGGCTTCCAGGAAACAAGGGGACATGGGCATGAAGGGAGAGACACGGGG 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 364 CCCAGTGGGAGCCAGGCCCCCAGGCCCCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 423
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 CCCATGGGGTCCCTGGAGCTCAGGGAGGTAAGGTGATGTGTAAGGAGAGAGAGAGAGAG 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 424 GGCCCTCCTGGCTTCCAGGAGTTCAGGAGCACCCTTTGGGGAGCCTTGGGTGCTTGACCT 483
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 GGTGGTGGTATCTCCAGAGGTCAAGGTGACCAAGGAAACCTGGAGTGCAGGGTGT 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 484 CGGGAGTGCACAGGCTTGCCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCCCCGGC 543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 CCAGGCCCTCAAGGTGACAGGAGCTTTCAGGTGCCAAGGAGTGCAGGAGGAGAGAGAGAG 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 544 CTTCTGGCCCATCAGGAGCGGTGG 568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 CTTCTGGCCAGCAGGACCCCCCG 1165
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-08-893-467A-1
; Sequence 1, Application US/08893467A
; Patent No. 6063901
; GENERAL INFORMATION:
; APPLICANT: Tryggevason, Karl
; APPLICANT: Elomaa, Outi
; APPLICANT: Kangas, Maarit
; TITLE OF INVENTION: An Isolated DNA Sequence For a
; Patent No. 6063901
; TITLE OF INVENTION: No. 6063901el Macrophage Receptor with
; TITLE OF INVENTION: a Collagenous Domain and the
; TITLE OF INVENTION: Polypeptide Chain Encoded by
; TITLE OF INVENTION: such a Sequence
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
; ADDRESSEE: Minnich & McKee
; STREET: 1100 Superior Avenue
; STREET: Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2518
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch,
; MEDIUM TYPE: 720 Kb storable
; COMPUTER: IBM PS/2, Model 35 SX
; OPERATING SYSTEM: DOS 5.0
```

```
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,467A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Minnich, Richard J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: TRV 2 009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 861-5582
; TELEFAX: (216) 241-1666
; TELEX: (216) 980162 \
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1868 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Nucleotide-genomic DNA
; HYPOTHETICAL: No. 6063901 relevant
; ANTI-SENSE: No. 6063901 relevant
US-08-893-467A-1

Query Match          9.7%; Score 99.4; DB 3; Length 1868;
Best Local Similarity 51.5%; Pred. No. 2.6e-17;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 124 CAAGGTCCACCGGGCCCCCAGAGGTCCAAAGGTGACAGAGATCCACAGGAGCCCTCGC 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 CCAGGTGTCCAAGGACCCAGGGCCCCCAGGACCAAGGAGGAGGAGGCTCCAGGGA 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 184 CCAACTGGCAACAAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 CTTAGGGTGCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 244 GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATCT 303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 GGCAGCAAGAGTGACATAGGTCTCACTGGCCCCCAAGGGGAAACATGGCACCAAGGAGAC 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 304 AAAGGCTCCAGGCCCCCAAGAGTCCCGTGGTCCCTGGGAGAGCCCGCCCTCAGGGC 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 AAAGGGACCTAGGCTTCCAGGAAACAAGGGGACATGGGCATGAAGGGAGAGACACGGGG 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 364 CCCAGTGGGAGCCAGGCCCCCAGGCCCCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 423
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 CCCATGGGGTCCCTGGAGCTCAGGGAGGTAAGGTGATGTGTAAGGAGAGAGAGAGAGAG 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 424 GGCCCTCCTGGCTTCCAGGAGTTCAGGAGCACCCTTTGGGGAGCCTTGGGTGCTTGACCT 483
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 GGTGGTGGTATCTCCAGAGGTCAAGGTGACCAAGGAAACCTGGAGTGCAGGGTGT 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 484 CGGGAGTGCACAGGCTTGCCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCCCCGGC 543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 CCAGGCCCTCAAGGTGACAGGAGCTTTCAGGTGCCAAGGAGTGCAGGAGGAGAGAGAGAG 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 544 CTTCTGGCCCATCAGGAGCGGTGG 568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 CTTCTGGCCAGCAGGACCCCCCG 1165
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
US-08-555-669-11
; Sequence 11, Application US/08555669
; Patent No. 5773248
; GENERAL INFORMATION:
; APPLICANT: Brewton, Richard G.
; APPLICANT: Mayne, Richard
; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
```

; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555.669
; FILING DATE: 13-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25/227
; REFERENCE/DOCKET NUMBER: 8389-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-854-3660
; TELEFAX: 415-854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..2098
; US-08-555-669-11

Query Match 9.7%; Score 99.4; DB 1; Length 2543;
Best Local Similarity 51.5%; Pred. No. 3e-17;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 124 CAAGGTCCACCGGGCCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCAGGGACCCCTGGC 183
DB 584 CCAGGTCCCGGAGGAGTCCGGAATGCCAGGGTTCAGAGGACCCACTGGCTACAAAGGC 643
QY 184 CCAACTGGCAACAAGGGACAGAAAGAGAGAGAGGGGAGCCCTGGACACCTGGCCCTGGG 243
DB 644 GAGCAGGGGGAAGTCCGGAAGGACGGCGGAGAGGCTGACCTGGCCCTGGCCCGCCG 703
QY 244 GGTGAGAGAGGCCCAATTTGGACACAGTGTCCCGCGGAGAGCGTGGCGGCAAAAGGATCT 303
DB 704 GCCTCCCGGGCAGCGTGGGGTCCAGGGCCCCCGGGGATTCAGAGGACTGCCAGGCCA 763
QY 304 AAAGGCTCCAGGGCCCCCAGGGTCCCGTGGTTCCTCCCTGGGAAGCCCGGCCCTCAGGC 363
DB 764 CTCGGGCCCCCTGGGGACCGGGGTCCCATTTGGTTCCGAGGGCCCCCTGGGATCCAGGA 823
QY 364 CCCAGTGGGACCCAGGGCCCCCGGGCCACAGGCAAGAGGGGACTCCCGGCCCTCAG 423
DB 824 GCGCTCCGGAAGCGGGGTACCGAGGCGAGAGGGGCCCAAGAGGTTCCGCGGCCCAAG 883
QY 424 GCGCTCTCTGGCTTCAGGGACTTCAGGGCACCGTTGGGGAGCCCTGGGGTGCCTGGACCT 483
DB 884 GTGACCTCGGACAGCTGGTCCCAAGGGAACCCCGGAGTGGCGGGCCAGCGGAGAG 943
QY 484 CCGGAGCTGCCAGGTTCCTCGGGGTACCGACATGCCAGGCCCCCAAGGGCCCCCGCCG 543
DB 944 CCGGGCATGCCAGGAAGGACGGCCAGAAATGCGTGCACGAGCTCGATGGCCAGAAGGA 1003
QY 544 CTCCTGCCCCATCAGGACGGGTGG 568
DB 1004 GAGGCTGGTTCGCAACGGTGTCTCCGG 1028

RESULT 14
US-09-073-663-11
; Sequence 11, Application US/09073663

; Patent No. 6127523
; GENERAL INFORMATION:
; APPLICANT: Brewton, Richard G.
; APPLICANT: Mayne, Richard
; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McGregor & Adler, LLP
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS 8.1
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073.663
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benjamin Aaron Adler, Ph.D., J.D.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5913D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 777-2321
; TELEFAX: (713) 777-6908
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..2098
; US-09-073-663-11

Query Match 9.7%; Score 99.4; DB 3; Length 2543;
Best Local Similarity 51.5%; Pred. No. 3e-17;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 124 CAAGGTCCACCGGGCCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCAGGGACCCCTGGC 183
DB 584 CCAGGTCCCGGAGGAGTCCGGAATGCCAGGGTTCAGAGGACCCACTGGCTACAAAGGC 643
QY 184 CCAACTGGCAACAAGGGACAGAAAGAGAGAGAGGGGAGCCCTGGACACCTGGCCCTGGG 243
DB 644 GAGCAGGGGGAAGTCCGGAAGGACGGCGGAGAGGCTGACCTGGCCCTGGCCCGCCG 703
QY 244 GGTGAGAGAGGCCCAATTTGGACACAGTGTCCCGCGGAGAGCGTGGCGGCAAAAGGATCT 303
DB 704 GCCTCCCGGGCAGCGTGGGGTCCAGGGCCCCCGGGGATTCAGAGGACTGCCAGGCCA 763
QY 304 AAAGGCTCCAGGGCCCCCAGGGTCCCGTGGTTCCTCCCTGGGAAGCCCGGCCCTCAGGC 363
DB 764 CTCGGGCCCCCTGGGGACCGGGGTCCCATTTGGTTCCGAGGGCCCCCTGGGATCCAGGA 823
QY 364 CCCAGTGGGACCCAGGGCCCCCGGGCCACAGGCAAAAGAGGGGACTCCCGGCCCTCAG 423
DB 824 GCGCTCCGGAAGCGGGGTACCGAGGCGAGAGGGGCCCAAGAGGTTCCGCGGCCCAAG 883
QY 424 GCGCTCTCTGGCTTCAGGGACTTCAGGGCACCGTTGGGGAGCCCTGGGGTGCCTGGACCT 483
DB 884 GTGACCTCGGACAGCTGGTCCCAAGGGAACCCCGGAGTGGCGGGCCAGCGGAGAG 943
QY 484 CCGGAGCTGCCAGGTTCCTCGGGGTACCGACATGCCAGGCCCCCAAGGGCCCCCGCCG 543
DB 944 CCGGGCATGCCAGGAAGGACGGCCAGAAATGCGTGCACGAGCTCGATGGCCAGAAGGA 1003

QY 544 CCTCCTGGCCCATCAGGAGGGTGG 568
|||||
Db 1004 GAGGCTGGTCGCAACGCTGCTCGG 1028

RESULT 15

US-09-484-970B-4
; Sequence 4, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 4359
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 331365 (1553795CB1)
US-09-484-970B-4

Query Match 9.5%; Score 97.6; DB 4; Length 4359;
Best Local Similarity 52.3%; Pred. No. 1.2e-16;
Matches 272; Conservative 0; Mismatches 234; Indels 14; Gaps 2;

QY 55 GTGATTATGGAGAAATGAAGCTAGTAGACTCAAGCATGGTCAGCTCATCAAGAATTTT 114
|||
Db 509 GAGCTGATGGAGGACCAAGGTCAGGGGCTGCCAGGAGACGAGTCTGTAAGGTTCC 568
|||
QY 115 ACAATACTACAGGTCCACCGGCCCCAGGGTCCAAAGAGTGACAGAGGATCCC----- 169
|||
Db 569 CAGGACCCCGAGGTTTCATAGGACCCGAGGATCCAAAGGTGCAGTGGGCTCCCTGGCC 628
|||
QY 170 ---AGGGACCCCTGCCCACTGGCAACAGGGACAGAGGAGAGAGAGGGGGGAGCCTG 226
|||
Db 629 CAGATGGATCCCGAGTCCCATCGGCTGCCAGGGCCAGATGGGCCCTCGGGGAAAGG 688
|||
QY 227 GACCACTGG-----CCCTGGGGTGAGAGAGGCCCAATGGACAGCTGGTCCGCCCG 280
|||
Db 689 GCCTCCTGGAGAGTCTCTGGGAGCTCAGCCGGGCCACGGGAGATGCTGGTGCCTG 748
|||
QY 281 GAGAGCTGGGGCAAGGATCTAAAGGTCCAGGGCCCCAAAGGCTCCCGTGGTTCCC 340
|||
Db 749 GACAGCTGGGCTTAAAGGCTTCCCGGAGACAGAGGCCCCCTGGATTTCAGAGGAAGC 808
|||
QY 341 CTGGGAAGCCCGCCCTCAGGGCCCCAGTGGGGACCCAGGGCCCCCGGGCCACAGGCA 400
|||
Db 809 AAGGGATGCTGGGATGCGCAGGCTGAAGGGCCAGCCAGGACCTCCAGGACCTTCGGCC 868
|||
QY 401 AGAGGGACTCCCGCCCTCAGGGCCCTCTGGGCTTCAGGGACCTTCAGGGCACCGTTG 460
|||
Db 869 AGCCAGGCTGTATGGGCTCCAGGACTGCATGGATTCCAGGAGCTCTGGCCAGAGG 928
|||
QY 461 GAGAGCTGGGGTCTGACCTCGGGGACTGCCAGGCTTGCTGGGGTACCAAGGATGC 520
|||
Db 929 GGCCCTTGGGGTCCAGGAATCCAGGCGGTGAAGGTCTGCTGTGATAGAGGGGACC 988
|||
QY 521 CAGGCCCCAAGGGCCCCCGGGCTCTTGGCCCATCAGG 560
|||
Db 989 CTGGGACACAGGCGCTCTGGCCCTGTGGCATGAAGG 1028
|||

Search completed: March 21, 2003, 08:25:53
Job time : 48.1652 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:57:56 ; Search time 156.339 Seconds
(without alignments)
5089.177 Million cell updates/sec

Title: US-09-763-712A-1_COPY_670_1695

Perfect score: 1026

Sequence: 1 atgcaacaagatttgatgag.....cagtactgtcatctgcatta 1026

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 38737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 1024.4 | 99.8 | 2930 | 10 | US-09-745-763-198 |
| 2 | 111 | 10.8 | 6728 | 10 | US-09-954-456-782 |
| 3 | 111 | 10.8 | 6728 | 10 | US-09-880-107-3946 |
| 4 | 110.6 | 10.8 | 1133 | 9 | US-09-924-340-57 |
| 5 | 110.6 | 10.8 | 1133 | 9 | US-09-992-600A-57 |
| 6 | 109.4 | 10.7 | 2691 | 10 | US-09-925-302-64 |
| 7 | 107.2 | 10.4 | 2192 | 10 | US-09-925-301-42 |
| 8 | 107 | 10.4 | 5086 | 10 | US-09-880-107-3947 |
| 9 | 107 | 10.4 | 5145 | 9 | US-09-925-299-206 |
| 10 | 107 | 10.4 | 5145 | 10 | US-09-925-299-206 |
| 11 | 107 | 10.4 | 5432 | 12 | US-10-044-090-22 |
| 12 | 106.8 | 10.4 | 4908 | 9 | US-10-001-887-33 |
| 13 | 106 | 10.3 | 2542 | 9 | US-09-954-531-961 |
| 14 | 106 | 10.3 | 2542 | 10 | US-09-964-824A-255 |
| 15 | 105.4 | 10.3 | 5416 | 10 | US-09-954-456-786 |
| 16 | 105.4 | 10.3 | 5416 | 10 | US-09-880-107-2094 |
| 17 | 105.2 | 10.3 | 6158 | 10 | US-09-919-497-6 |
| 18 | 105.2 | 10.3 | 6158 | 10 | US-09-954-456-762 |
| 19 | 104.2 | 10.2 | 4149 | 12 | US-10-044-090-104 |

| | | | | | |
|----|-------|------|------|----|--------------------|
| 20 | 102.4 | 10.0 | 1486 | 10 | US-09-925-302-247 |
| 21 | 101.6 | 9.9 | 3226 | 10 | US-09-954-456-725 |
| 22 | 101.4 | 9.9 | 3690 | 12 | US-10-044-090-448 |
| 23 | 98.6 | 9.6 | 2158 | 9 | US-10-001-835-98 |
| 24 | 97.8 | 9.5 | 2520 | 10 | US-09-880-107-3685 |
| 25 | 95.8 | 9.3 | 1485 | 10 | US-09-925-302-246 |
| 26 | 95.2 | 9.3 | 1797 | 9 | US-09-978-295A-613 |
| 27 | 95.2 | 9.3 | 1797 | 9 | US-09-978-697-613 |
| 28 | 95.2 | 9.3 | 1797 | 9 | US-09-978-192A-613 |
| 29 | 95.2 | 9.3 | 1797 | 9 | US-09-999-832A-613 |
| 30 | 95.2 | 9.3 | 1797 | 9 | US-09-978-189-613 |
| 31 | 95.2 | 9.3 | 1797 | 9 | US-10-174-590-331 |
| 32 | 95.2 | 9.3 | 1797 | 9 | US-10-176-758-331 |
| 33 | 95.2 | 9.3 | 1797 | 9 | US-10-175-737-331 |
| 34 | 95.2 | 9.3 | 1797 | 9 | US-10-173-706-331 |
| 35 | 95.2 | 9.3 | 1797 | 9 | US-10-175-738-331 |
| 36 | 95.2 | 9.3 | 1797 | 9 | US-10-175-752-331 |
| 37 | 95.2 | 9.3 | 1797 | 9 | US-10-176-482-331 |
| 38 | 95.2 | 9.3 | 1797 | 9 | US-10-176-757-331 |
| 39 | 95.2 | 9.3 | 1797 | 9 | US-10-176-913-331 |
| 40 | 95.2 | 9.3 | 1797 | 9 | US-10-180-552-331 |
| 41 | 95.2 | 9.3 | 1797 | 9 | US-10-180-557-331 |
| 42 | 95.2 | 9.3 | 1797 | 9 | US-10-173-700-331 |
| 43 | 95.2 | 9.3 | 1797 | 9 | US-10-174-572-331 |
| 44 | 95.2 | 9.3 | 1797 | 9 | US-10-174-579-331 |
| 45 | 95.2 | 9.3 | 1797 | 9 | US-10-174-582-331 |

ALIGNMENTS

RESULT 1

US-09-745-763-198
; Sequence 198 Application US/09745763
; Patent No. US20020065394A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth
McCoy, John M.
Lavallie, Edward R.
Collins-Racie, Lisa A.
Evans, Cheryl
Merberg, David
Treacy, Maurice
Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION NUMBER:

APPLICATION NUMBER: US/09/745,763

FILING DATE: 18-Jun-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.

REGISTRATION NUMBER: 41,323

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8284

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 198:

SEQUENCE CHARACTERISTICS:

LENGTH: 2930 base pairs

TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-09-745-763-198

Query Match 99.8%; Score 1024.4; DB 10; Length 2930;
Best Local Similarity 99.9%; Pred. No. 1.7e-271;
Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAACAGATTGATGAGGTCGAGGTAGACACTGAAGTACGCCAAGTATCACTGATT 60
DB 1266 ATGCAACAGATTGATGAGGTCGAGGTAGACACTGAAGTACGCCAAGTATCACTGATT 1325
QY 61 ATGGAAGAAATGAAGTAGTAGACTCCAAAGCATGTGTCAGCTCATCAAGAAATTTTACAATA 120
DB 1326 ATGGAAGAAATGAAGTAGTAGACTCCAAAGCATGTGTCAGCTCATCAAGAAATTTTACAATA 1385
QY 121 CTAAAGAGTCCACGGGCCCCAGGGTCCAAGAGTGACAGAGGATCCAGAGGACCCCT 180
DB 1386 CTAAAGAGTCCACGGGCCCCAGGGTCCAAGAGTGACAGAGGATCCAGAGGACCCCT 1445
QY 181 GCCCAACTGGCAACAAGGACAGAAAGAGAGAGAGGGGAGCCCTGGACCACTGGCCCT 240
DB 1446 GCCCAACTGGCAACAAGGACAGAAAGAGAGAGGGGAGCCCTGGACCACTGGCCCT 1505
QY 241 GCGGGTAGAGAGGCCCCAATTTGACAGCTGTGTCCTCCCGGAGAGCGTGGCGCAAGGA 300
DB 1506 GCGGGTAGAGAGGCCCCAATTTGACAGCTGTGTCCTCCCGGAGAGCGTGGCGCAAGGA 1565
QY 301 TCTAAGAGTCCACGGGCCCCAAGGCTCCCTGGTTCCTTGGGAGCCGCGCCCTCAG 360
DB 1566 TCTAAGAGTCCACGGGCCCCAAGGCTCCCTGGTTCCTTGGGAGCCGCGCCCTCAG 1625
QY 361 GCGCCAGTGGGAGCCAGGCCCCCGGCCACCAAGGAGAGGAGTCCCGGCCCT 420
DB 1626 GCGCCAGTGGGAGCCAGGCCCCCGGCCACCAAGGAGAGGAGTCCCGGCCCT 1685
QY 421 CAGGCCCCCTCTGGCTTCAGGAGCTTCAGGCAACCGTTGGGAGCGTGGGTGCTGA 480
DB 1686 CAGGCCCCCTCTGGCTTCAGGAGCTTCAGGCAACCGTTGGGAGCGTGGGTGCTGA 1745
QY 481 CTTGGGAGTCCAGGCTTGGGTGACAGGATGCCAGGCCCCAAGGCCCCCCCC 540
DB 1746 CTTGGGAGTCCAGGCTTGGGTGACAGGATGCCAGGCCCCAAGGCCCCCCCC 1805
QY 541 GCGCTCTCTGGCCATCAGAGCGGTGTGTCCTGGCCCTCAGAAATGAGCAACCCCG 600
DB 1806 GCGCTCTCTGGCCATCAGAGCGGTGTGTCCTGGCCCTCAGAAATGAGCAACCCCG 1865
QY 601 GCACGGAGGACAATGGCTGCGCCCTCAGTGAAGAACTTCACAGACAAATGCTACTAT 660
DB 1866 GCACGGAGGACAATGGCTGCGCCCTCAGTGAAGAACTTCACAGACAAATGCTACTAT 1925
QY 661 TTTTCAGTTGAGAAAGAAATTTTGAAGATGCAAGCTTTTCTGTGAGACAAAGTCTCA 720
DB 1926 TTTTCAGTTGAGAAAGAAATTTTGAAGATGCAAGCTTTTCTGTGAGACAAAGTCTCA 1985
QY 721 CATCTTTGTTTCAATAAAGTACAGAGAACAGCAATGATGATAAAGAACAGATGTAGGG 780
DB 1986 CATCTTTGTTTCAATAAAGTACAGAGAACAGCAATGATGATAAAGAACAGATGTAGGG 2045
QY 781 AGAGAGGCCACTGATCGGCTCACAGCTCAGAGCGTGAAGTGAATGAAAGTGGCTG 840
DB 2046 AGAGAGGCCACTGATCGGCTCACAGCTCAGAGCGTGAAGTGAATGAAAGTGGCTG 2105
QY 841 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACGCCGATTAAGTGGGTCAT 900
DB 2106 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACGCCGATTAAGTGGGTCAT 2165
QY 901 GCGCATGGGCCAGGAGAGACTGTCTGGTTGATTATGCTGGGAGTGGAAACGATTTC 960
DB 2166 GCGCATGGGCCAGGAGAGACTGTCTGGTTGATTATGCTGGGAGTGGAAACGATTTC 2225

QY 961 CAATGTGAAGACGTCAATAAATTCATTTCCGAAAAAGACAGGAGACAGTACTGTCTATCT 1020
DB 2226 CAATGTGAAGACGTCAATAAATTCATTTCCGAAAAAGACAGGAGACAGTACTGTCTATCT 2285
QY 1021 GCATTA 1026
DB 2286 GCATTA 2291

RESULT 2

US-09-954-456-782
; Sequence 782, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 782
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-782

Query Match 10.8%; Score 111; DB 10; Length 6728;
Best Local Similarity 53.0%; Pred. No. 3.7e-20;
Matches 237; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 122 TACAGGTCCACCGGCCCCAGGGTCCCAAGAGGTGACAGAGATCCACAGGACCCCTG 181
DB 721 TCCAAAGTCCCGCTGGTGAGCTGCGAGCTTCAGGTCCCATGGTCCCGGAG 780
QY 182 GCCCAACTGGCAACAAGGACAGAAAGAGAGAGGAGGAGCTGGACCACTGGCCCTG 241
DB 781 GTCCCCCAGGTCCCGCTGGAAGAAATGAGATGATGGGGAAGCTGGAACCTGGTCGTC 840
QY 242 CGGGTGAAGAGGCCAATTTGGACAGCTGTGTCCTCCCGGAGAGCGTGGCGCAAGGAT 301
DB 841 CTGTGTGAGCGTGGGCTCTCTGGGCTCAGGTGCTCGAGGATTTGCCCGGAACAGCTGGCC 900
QY 302 CTAAGAGCTCCAGGCCCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGG 361
DB 901 TCCCTGGAATGAAGGACACAGAGGTTTCAGTGGTTTGGATGGTGCCAAAGAGATGCTG 960
QY 362 GCGCCAGTGGGACCCAGGCCCCCGGGCCACACAGGCAAGAGGAGTCCCGGCGCTC 421
DB 961 GTCTGTGCTTAAAGGTGAGCTGGCAGCCCTGGTGAAATGGAGTCTCTGCTCAGA 1020

| | | | | |
|-----------------------|--------------|--|-----------------|-------------------|
| Query Match | 10.88; | Score 111; | DB 10; | Length 6728; |
| Best Local Similarity | 53.08; | Pred. No. 3,7e+20; | | |
| Matches 237; | Conservative | 0; | Mismatches 210; | Indels 0; Gaps 0; |
| Qy | 122 | TACAAGGTCACACGGGGCCCCAGAGGGTCCAAGAGGTACAGAGAGATCCACAGGGACCCCGCTG | 181 | |
| Db | 721 | TCCAAGGTCCCCCTGGTTCAGCCTGGCGAGCGCTGAGCTTCAGGTCCCATGGGTCCCGCAG | 780 | |
| Qy | 182 | GCCCACTGGCAACAAGGGACAGAGAAGAGAGGGGAGCGCTGGACCACCTGGCCCGCTG | 241 | |
| Db | 781 | GTCCCCCAGGTCCCGCTTGGAAAGAATGGAGATGATGGGGAAGCTTGGAAAACCTGTCGTCTC | 840 | |
| Qy | 242 | CGGCTGAGAGAGGCCCAATTTGGACACGTGTGTCCTCCCGGAGAGCGTGGCGCAAAAGGAT | 301 | |
| Db | 841 | CTGCTGAGCTGGGCCCTCTGGGCCCTCAGGTGCTCGAGGATTCGCCGGAACACGTGGCC | 900 | |
| Qy | 302 | CTAAAGGCTCCCAAGGGCCCCAAAAGCTCCCGTGGTTCCTCGTGGAAAGCCCGGCCCTCAGG | 361 | |
| Db | 901 | TCCCTTGAATGAAGGGACACAGAGGTTTCAGTGGTTTGGATGGTCCCAAGGAGATGCTG | 960 | |
| Qy | 362 | GCCCCAGTGGGGACCCAGGCCCCCGGGCCCCACAGACGAAGAGGACTCCCGCGGCCCTC | 421 | |
| Db | 961 | GTCTCTGTGTCTTAAGGGTGAGCCTTGCACGCCCTGTGTGAAANTGGAGCTCCTGGTCAGA | 1020 | |
| Qy | 422 | AGGCCCTCTCTGGCTTCCAGGGACTTCAGGCAACCGTTTGGGGAGCCTTGGGTGCTCGGAC | 481 | |
| Db | 1021 | TGGGCCCCGTGGCCTGCCTGGTGAGAGAGTCCGCCCTGGAGCCCTTGGCCCTGCTGGTG | 1080 | |
| Qy | 482 | CTCGGGGACTCCAGAGCTTGCCTGGGTACAGGACATCCAGGCCCCCAAGAGGGCCCCCCC | 541 | |
| Db | 1081 | CTCGTGAATATGATGGTGTCTACTGGTGTCTGCGGGGCCCCCTGGTGTCCACCGCCCCCGCTG | 1140 | |
| Qy | 542 | GCCCTCTCTGGCCCCATCAGAGCGGTGG | 568 | |

Db 211 CAMGGCGTCCGGGAGCTTAAAGGTCTCCCGGGTCTCTCTGCGCCGCCACAGGTCTTACGAC 270
Qy 362 -GCCCCAGTGGGACCCAGGCCCGCCCGGACCAAGAGGAGGACTCCCGGGCCCT 420
Db 271 ATCATCAAGGGAGCGCCGGCTCCTGGTCTCTGAGGCCGCCACAGGCTCAAGAGGGCTT 330
Qy 421 CAGGCGCTCTCTGGCTTCCAGGGACTTACAGGCACCGCTTGGGAGCCCTGGGGTCCCTGGA 480
Db 331 CAGGAGCTGCCAGGCCCGGAAAGGCCAGCAAGGTGTACAGGATTGGTGGGTATACCTGGA 390
Qy 481 CCGGGGACTGCCAGGCTTCCCTGGGTACAGGCATGCCAGGCCAGGCCCGCCCGCC 540
Db 391 CCGCAGGTATCTCTGGGTGTGACGGTTCCTCGCCAGAAAGAGAGATGGGACCTGCC 450
Qy 541 GGCCTCTCTGCCCATCAGGA 561
Db 451 GGGCTACTGTCCAAGGA 471

RESULT 7
US-09-925-301-42
; Sequence 42, Application US/09925301
; Patent No. US2002052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 2192
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-301-42

Query Match 10.4%; Score 107.2; DB 10; Length 2192;
Best Local Similarity 53.5%; Pred. No. 2.2e-19;
Matches 274; Conservative 0; Mismatches 228; Indels 10; Gaps 2;

Qy 127 GGTCCACCGGCCCCAGGGTCCAAAGGTGACAGAGATCCAGGACCCCTGGGCCA 186
Db 638 GCGCCCGGTGACCCCTGGTGTCTCTGGTCTCTGTGCGCCCTGGCCCGTTGGCCCT 697
Qy 187 ACTGGCAACAGGACAGAAAGGAGAGAAAGGGGAGGCTGGACACCTGGCCCTGCGGGT 246
Db 698 GCTGGCAAGAGTGTGATCGTGTGAGACTGTCTCTGTGTCGCGCGGTCTGTGCGC 757
Qy 247 GAGAGAGCCCAATTTGACACAGCTGTGTCGCCCGGAGAGCGTGGCGCAAGAGTCTAAA 306
Db 758 CCTGTTGGCGCCCTGGCGCCCGCGGACCCCAAGGCCCGCTGGTGACAAGGGTGAGACA 817
Qy 307 GCGTCCAGGCGCCCAAGGCTCCGTTGTTCCCTGGAGAGCCCGCCCTCAGGGCCCG 366
Db 818 GCGCAAGAGGCGACAGAGGTCATAAGGGTCAACGTTCTCTGCGCTTCAGGGTCC- 876
Qy 367 AGTGGGACCCAGGCCCGCCCGGCGCCACAGGCAAGAGGAGTCCCGGCCCTCAGGGC 426
Db 877 -----CCCTGGCCCTCTGGCTCTCTGTTGTAACAAAGTCCCTGTGAGACCTCTGGT 928
Qy 427 CCTCCTGGCTTCCAGGACTTACAGGACCGCTTGGGAGCGTGGGGTGCCTTGGACCTCGG 486
Db 929 CCTGCTGTCTCCCGAGGTCCCGCTGGCTCTGCTGGTCTCTGCTGCGCAAGAGTGAAC 988
Qy 487 GGACTGCCAGCTTGCCTGGGTACAGGATGTCAGGCCCGCAAGGCCCGCCCGCCCT 546
Db 989 GGTCTCCTGGCCCAATTGGGCGCCCTGGTCTCTGCGGTGCGACTGGTGTGCTGCTCT 1048

Qy 547 CTTGGCCCATCAGGAGCGG-TGGTCCCTTGCCCTGAGATGAGCAACCCCGGACC 605
Db 1049 GTTGTCCCGCCCGGCTCTCTGGACCTCTGTGTCCTCCCTGCTCCACAGCGCTGTTTC 1108
Qy 606 GGAGGACAATGGTGGCGGCTCACTGGAAGA 637
Db 1109 GACTTTCAGCTTCTCTGCGCCCGGACCTCAAGA 1140

RESULT 8
US-09-880-107-3947
; Sequence 3947, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3947
; LENGTH: 5086
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 Z74616
; NAME/KEY: unsure
; LOCATION: (1)..(5086)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3947

Query Match 10.4%; Score 107; DB 10; Length 5086;
Best Local Similarity 51.3%; Pred. No. 4e-19;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

Qy 120 ACTCAAGGTCCACCGGCCCGCCCAAGAGGTGACAGAGATCCAGGACCCCG 179
Db 709 ACAGCCCGGTCTCTCTGTGTGAAGGTGAACCTGGTCCCTGGTGAATGGAATCC 768
Qy 180 TGGCCCAACTGGCAACAGGACAGAAAGAGAGAGAGGGGAGCGCTGGACACCTGGGCC 239
Db 769 AGGTCAACAGGAGCGCCGTGGCTTCTGTGTGAGAGAGAGCGTGTGGTGGCCCTGGGCC 828
Qy 240 TCGGGGTGAGAGAGGCCCAATTGGACAGCTGTGTCGCCCGGAGAGCGTGGCGGCAAGG 299
Db 829 AGCTGTGTCCCGTGGCATGGAAGTGTGGGTCCCGTGGGTCTCTGCTGTCCCATTTG 888
Qy 300 ATCTAAAGGTCCACAGGCGCCCAAGGCTCCCGTGGTTTCCCTTGGAGAGCCCGCCCTCA 359
Db 889 GTCTGTGGCCCTCCAGGCTTCCAGGTTCCCGCCCTGGCCCAAGGGTGAATTTGAGCTGT 948
Qy 360 GGGCCCGAGTGGGAGCCCGAGGCCCGCCCGGCGCCACAGGCAAGAGGAGTCCCGCGGCC 419
Db 949 TGGTAACGCTGTCTCTGTGTCCCGCGGTCCCGTGGTGAAGTGGTCTTCCAGGCT 1008
Qy 420 TCAGGGCCCTCTGTGTCTCAGGAGCTTTCAGGGACCGCTTGGGAGCGCTGGGGTGCCTGG 479
Db 1009 CTCGGGCGCCCTTGGACCTCTGTGTAATCTTGAGCAAAAGCGCTTACTGTGCAAGGG 1068
Qy 480 ACCTCGGGAGTGCAGGCTTGCCTGGGTACAGGATGCCAGGCCCGCCCGGAGGCGCCCG 539
Db 1069 TGCTGTGGCCCTTCCCGGCGTGTGCTGGGCTCCCGGCGCTCTCTGGACCCCGGCTATTC 1128
Qy 540 CGGCCCTCTGGCCCATCAGGAGCGGTGGTGGCCCTTGGCCCTTGCAAGATGAGCAACCC 599

Db 1129 TGGCCCTGTTGGTGGCGGTGCTACTGGTCCAGAGACTTGTGGTGAGCCTGGTCC 1188
QY 600 GGC 602
Db 1189 AGC 1191

RESULT 9
US-09-925-299-206
; Sequence 206, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Query Match 10.4%; Score 107; DB 9; Length 5145;
Best Local Similarity 51.3%; Pred. No. 4e-19;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;
QY 120 ACTACAAGTCCACCGGGCCCGAGGGTCCAGAGGTGACAGAGATCCCGAGGGACCC 179
Db 712 ACAGCCGGTGTCTTGGTGTGAAGGGTGAACCTGGTCCCTTGGTGAAGTGAAGTGA 771
QY 180 TGGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGAGGGGAGCGCTGGACCACTGGCC 239
Db 772 AGGTCAACAGAGCCGTGGGCTTCTGTGTGAGAGAGACGTGTGGTCCCTGGCC 831
QY 240 TGGCGGTGAGAGAGGCCCAATTTGACAGCTGGTCCCGCCGAGAGAGCGTGGCGGCAAGG 299
Db 832 AGCTGGTCCGCTGGCAGTGTGAAGTGTGGGTCCCGTGGGTCTCTGTGGTCCCATTTGG 891
QY 300 ATCTAAGGCTCCAGGGCCCAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCCGGCCCTCA 359
Db 892 GTCTGTGCTCCAGGCTTCCAGGTTCCAGGTCCTCCCGCCCAAGGGTGAATTTGGARCTGT 951
QY 360 GGGCCCACTGGGGACCCAGGGCCCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
Db 952 TGGTACGCTGGTCTGTGGTGGTCCCGCCGCTCCCGTGTGAAGTGGGTCTTCCAGGCT 1011
QY 420 TCAGGCGCTCTGTGGTCTCCAGGAGCTTCAGGCGACCGTGGGGAGCGCTGGGGTGCCTGG 479
Db 1012 CTCGGGCGCGTGGACCTTCCTGGTAACTCTGGAGCAACGGCTTACTGGTGCCAAGG 1071
QY 480 ACCTCGGGGAGTCCAGGCTTGCCTGGGTACCGAGCATGCCAGGCCCAAGGGCCCCC 539

Db 1072 TGCTGTGGCCTTCCCGCGTGTGGGGCTCCCGGCTCCCTGGACCCCGGGTATTCC 1131
QY 540 CGGCCTCTCTGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCACCC 599
Db 1132 TGGCCCTGTTGGTGTGGCGGTGCTACTGGTCCAGAGACTTGTGGTGAGCCTGGTCC 1191
QY 600 GGC 602
Db 1192 AGC 1194

RESULT 10
US-09-925-299-206
; Sequence 206, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Query Match 10.4%; Score 107; DB 10; Length 5145;
Best Local Similarity 51.3%; Pred. No. 4e-19;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;
QY 120 ACTACAAGTCCACCGGGCCCGAGGGTCCAGAGGTGACAGAGATCCCGAGGGACCC 179
Db 712 ACAGCCGGTGTCTTGGTGTGAAGGGTGAACCTGGTCCCTTGGTGAAGTGAAGTGA 771
QY 180 TGGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGAGGGGAGCGCTGGACCACTGGCC 239
Db 772 AGGTCAACAGAGCCGTGGGCTTCTGTGTGAGAGAGACGTGTGGTCCCTGGCC 831
QY 240 TGGCGGTGAGAGAGGCCCAATTTGACAGCTGGTCCCGCCGAGAGAGCGTGGCGGCAAGG 299
Db 832 AGCTGTGCTCCAGGCTTCCAGGTTCCAGGTCCTCCCGCCCAAGGGTGAATTTGGARCTGT 891
QY 300 ATCTAAGGCTCCAGGGCCCAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCCGGCCCTCA 359
Db 892 GTCTGTGCTCCAGGCTTCCAGGTTCCAGGTCCTCCCGCCCAAGGGTGAATTTGGARCTGT 951
QY 360 GGGCCCACTGGGGACCCAGGGCCCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
Db 952 TGGTACGCTGGTCTGTGGTGGTCCCGCCGCTCCCGTGTGAAGTGGGTCTTCCAGGCT 1011
QY 420 TCAGGCGCTCTGTGGTCTCCAGGAGCTTCAGGCGACCGTGGGGAGCGCTGGGGTGCCTGG 479
Db 1012 CTCGGGCGCGTGGACCTTCCTGGTAACTCTGGAGCAACGGCTTACTGGTGCCAAGG 1071

```
QY 480 ACTCTGGGGAAGCTCCAGGCTTCCTGGGTACCAAGGATGCCAGGCCCCCAAGGGCCCCC 539
    || || || || || || || || || || || || || || || || || || || || ||
Db 1072 TGTCTGTGCTTCCCGGCTTCTGGGCTCCCGGCTCCCTGGACCCCGGGTATTCC 1131
    || || || || || || || || || || || || || || || || || || || || ||
QY 540 CGGCCCTCTGCCCCATCAGGAGGGTGTGCCCCCTGCGCCTGCAGATGACCAACCCC 599
    || || || || || || || || || || || || || || || || || || || || ||
Db 1132 TGGCCCTGTTGTGCTGCGGGTCTACTGGTCCAGAGGACTTGTGTGAGCCTGGTCC 1191
    || || || || || || || || || || || || || || || || || || || || ||
QY 600 GGC 602
    ||
Db 1192 AGC 1194

RESULT 11
US-10-044-090-22
; Sequence 22, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 5432
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.13
US-10-044-090-22

Query Match 10.4%; Score 107; DB 12; Length 5432;
Best Local Similarity 51.3%; Pred. No. 4.1e-19;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

QY 120 ACTAAGGTCTCCAGGGCCCCAGGGTCCAAAGAGGTGACAGGATCCAGGAGCCCCC 179
    || || || || || || || || || || || || || || || || || || || || ||
Db 1042 ACAGCCGGTGTCTCTGTGTGAAGGGTGAACCTGTGCCCCCTGGTGAATGGAACTCC 1101
    || || || || || || || || || || || || || || || || || || || || ||
QY 180 TGGCCCCAAGTGGGACCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 239
    || || || || || || || || || || || || || || || || || || || || ||
Db 1102 AGGTCAACAGAGAGCCCGTGGCTTCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1161
    || || || || || || || || || || || || || || || || || || || || ||
QY 240 TCGGGGTGAGAGAGGCCCAATTGGACAGCTGGTCCCCCGGAGAGAGCGTGGCGGCAAGG 299
    || || || || || || || || || || || || || || || || || || || || ||
Db 1162 AGCTGGTGGCCGTGGCAGTGTGGAAGTGTGGTCCCGTGGTCCCTGTGTGTCCTATTGG 1221
    || || || || || || || || || || || || || || || || || || || || ||
QY 300 ATCTAAGGCTCCAGGGCCCCAAGGCTCCCGTGGTTCCTCTGGAGAGCCCGGCCCTCA 359
    || || || || || || || || || || || || || || || || || || || || ||
Db 1222 GTCTGTGCCCCCTCAGGCTTCCAGGTGCCCTCGCCCTGGCCCCAAGGGTGAATTTGAGCTGT 1281
    || || || || || || || || || || || || || || || || || || || || ||
QY 360 GGGCCCCAGTGGGACCCAGGCCCCCGGGCCCCACAGCAAGAGGAGACTCCCCGGCCC 419
    || || || || || || || || || || || || || || || || || || || || ||
Db 1282 TGGTAAGCTGTCTCTGTGTGTCGCCCGGCTCCCGTGTGTAAGTGGGTCTTCCAGGCTT 1341
    || || || || || || || || || || || || || || || || || || || || ||
QY 420 TCAGGGCCCTCTGCTTCCAGGAGACTTCAGGGCACCCTGTGGGAGAGCTTGGGGTGCCTGG 479
    || || || || || || || || || || || || || || || || || || || || ||
Db 1342 CTCGGGCCCGTGGACCTCTCTGTAATCTTGAGCAACAGGCCCTTACTGTGTCGAAGG 1401
    || || || || || || || || || || || || || || || || || || || || ||
QY 480 ACCTCGGGAGCTGCAGAGCTTCCCTGGGTACCAAGGATGCCAGGCCCCCAAGGGCCCCC 539
    || || || || || || || || || || || || || || || || || || || || ||
Db 1402 TGTCTGTGCCCCCTCCGGGCTGTGCTGGGGCTCCCGGCTCCCTGGAGCCCCCGGCTATTCC 1461
    || || || || || || || || || || || || || || || || || || || || ||
QY 540 CGGCCCTCTGCCCCATCAGGAGCGGTGTGCCCCCTGCGCCTGCAGATGAGCCAAACCCC 599
    || || || || || || || || || || || || || || || || || || || || ||
Db 1462 TGGCCCTGTGTGTGCTGCGGGTGTCTACTGGTCCAGAGGACTTGTGTGTGAGCCTGGTCC 1521
    || || || || || || || || || || || || || || || || || || || || ||
QY 600 GGC 602
    ||
```

Db 1522 AGC 1524

RESULT 12

US-10-001-887-33

; Sequence 33, Application US/10001887

; Patent No. US2002015464A1

; GENERAL INFORMATION:

; APPLICANT: Salceda, Susana

; APPLICANT: Macina, Roberto

; APPLICANT: Recipon, Herbe

; APPLICANT: Cafferkey, Robert

; APPLICANT: Sun, Yongming

; APPLICANT: Liu, Chenghua

; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and

; FILE REFERENCE: DEX-0269

; CURRENT APPLICATION NUMBER: US/10/001,887

; CURRENT FILING DATE: 2001-11-20

; PRIOR APPLICATION NUMBER: 60/249,998

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: 60/252,563

; PRIOR FILING DATE: 2000-11-22

; NUMBER OF SEQ ID NOS: 137

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 33

; LENGTH: 4908

; TYPE: DNA

; ORGANISM: Homo sapien

US-10-001-887-33

Query Match

Best Local Similarity 10.4%; Score 106.8; DB 9; Length 4908;

Matches 256; Conservative 0; Mismatches 222; Indels 6; Gaps 1;

```
QY 126 AGGTCCACCGGGCCCCAGGGTCCAAAGAGTGACAGAGATCCAGGAGACCCCTGGCCC 185
    || || || || || || || || || || || || || || || || || || || || ||
Db 1603 AGGGCTGGAGGACCCCAAGAGTGCCTGGCATTCAGAGTGACCAAGGGCCTAGTGGCCT 1662
    || || || || || || || || || || || || || || || || || || || || ||
QY 186 AACTGSCAAAGGACACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 245
    || || || || || || || || || || || || || || || || || || || || ||
Db 1663 GGCTGGGAACACAGGGGTCCCAGGTGAGAGGGAGACTTCTGGGGCCATTTGGACCCCTGG 1722
    || || || || || || || || || || || || || || || || || || || || ||
QY 246 TGAGAGAGAGCCCAATTGGACAGCTGGTCCCCCGGAGAGAGAGAGAGAGAGAGAGAG 305
    || || || || || || || || || || || || || || || || || || || || ||
Db 1723 ACCAACTGGGCCCAAGGGTGAAGCGGTTTACGGGTCCCTGGAGGAGACAGGGGTGGC 1782
    || || || || || || || || || || || || || || || || || || || || ||
QY 306 AGGCTCCAGGAGGCCCAAGAGCTCCCGTGGTTCCTGGGAAAGCCCGGCCCTCAGGGCCCC 365
    || || || || || || || || || || || || || || || || || || || || ||
Db 1783 AGGAGCCCTGGGGCAGAAAGGTGACTTGGGGTCTCCCTGGGACAGCTTGGGCTGAGGGGTCC 1842
    || || || || || || || || || || || || || || || || || || || || ||
QY 366 CAGTGGGAGCCAGAGCCCGGGCCCCACAGGCAAGAGAGGAGACTCCCGGCCCTCAGGG 425
    || || || || || || || || || || || || || || || || || || || || ||
Db 1843 CTCAGAAATCCAGGACTCCAGGGTCCAGGTCCAGTGGCCCTATTGGGCCCAAGGGCTGCCGGG 1902
    || || || || || || || || || || || || || || || || || || || || ||
QY 426 CCTCTGCTTCTCCAGGGAATTCAGGGCACCGTTTGGGAG-----CTTGGGGTGCCTGG 479
    || || || || || || || || || || || || || || || || || || || || ||
Db 1903 CTTAAGGGGGAACAGGCTCCAGGGCCCCCTGGAGAGGGGAGAGAGAGGGAACCTGG 1962
    || || || || || || || || || || || || || || || || || || || || ||
QY 480 ACCTCGGGAGCTGCAGAGCTTGCCTGGGTACCAAGGATGCCAGGCCCCCAAGGGCCCCC 539
    || || || || || || || || || || || || || || || || || || || || ||
Db 1963 CACGGTGGGGCCAGGGGCCCCAGGGGTCCCTTGGCTCCCTTGAATACAGGGGCCCTCC 2022
    || || || || || || || || || || || || || || || || || || || || ||
QY 540 CGGCCCTCTGCCCCATCAGAGCGGTGTGCCCCCTGCGCCTGCAGAAATGAGCAACCCC 599
    || || || || || || || || || || || || || || || || || || || || ||
Db 2023 GGGGCTCCCGGGCCCCCGGAGACCCCTGTGCCCTTGGGGCTTCGATGAGACTGGCAT 2082
    || || || || || || || || || || || || || || || || || || || || ||
QY 600 GGC 603
    ||
Db 2083 GGC 2086

RESULT 13
US-09-954-531-961
```

; Sequence 961, Application US/09954531
 ; Patent No. US20020165180A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Weaver, Zoe
 ; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
 ; TITLE OF INVENTION: Gene Sets
 ; FILE REFERENCE: 689290-77
 ; CURRENT APPLICATION NUMBER: US/09/954,531
 ; CURRENT FILING DATE: 2002-05-02
 ; PRIOR APPLICATION NUMBER: US/60/233,133
 ; PRIOR FILING DATE: 2000-09-18
 ; PRIOR APPLICATION NUMBER: US/60/234,009
 ; PRIOR FILING DATE: 2000-09-20
 ; PRIOR APPLICATION NUMBER: US/60/234,034
 ; PRIOR FILING DATE: 2000-09-20
 ; PRIOR APPLICATION NUMBER: US/60/234,509
 ; PRIOR FILING DATE: 2000-09-22
 ; PRIOR APPLICATION NUMBER: US/60/234,567
 ; PRIOR FILING DATE: 2000-09-22
 ; NUMBER OF SEQ ID NOS: 1392
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 961
 ; LENGTH: 2542
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; ORGANISM: Homo sapiens
 US-09-954-531-961

Query Match 10.3%; Score 106; DB 9; Length 2542;
 Best Local Similarity 52.1%; Pred. No. 5 le-19;
 Matches 236; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

QY 122 TACAAGTCCACCGGGCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCAGGAGACCCCTG 181
 Db 630 TCCCTGACCAACCGGGTCCCAGGGCATCAGGGGCTACCCAGGCATGGCAGGGGCCAAGG 689
 QY 182 GCCCAACTGGCAACAGGGGACGAAGAAGAGAGAGAGGGGGAGCCCTGGACCACTGGCCCTG 241
 Db 690 GAGAGACGGGCGCCTCATGGATATAAAGGCATGTGTGGCGCTATCGGTGCCACTTGGGCCAC 749
 QY 242 CGGTGAGAGAGCGCCCAATTGGACACAGCTGTCCCGCCGGAGAGCGTGGCGGCAAAAGGAT 301
 Db 750 CGGTGAGGAAGGTCTTAGGGACCGCCAGCGAGCTGGGAGAGAGGTGTACAGAGGCCA 809
 QY 302 CTAAGGCTCCAGGGCCCCCAAGGCTCCGTGGTTCCTCTGGGAAGCCCGGCCCTCAGG 361
 Db 810 GCCCAGGTATTCTGTGGACCCAGGGGATCAGCGCCGAAAGAGCAACGGGCCCCACG 869
 QY 362 GCCCAGTGGGACCCAGGCCCCCGGGCCACCAAGAGAGGAGCTCCCGGGCCCTC 421
 Db 870 GCATCAACGGCAAGGATGGGACCCCGCCAGCGCTGGCATGAAGGGCAGTGCAGGACAGG 929
 QY 422 AGGCGCTCTCTGGCTTCCAGGGACTTCAGGSCACCGTTGGGAGCCTGGGTGCTGGAC 481
 Db 930 CGGACACCGCGGAAGTCCAGGCCACCAAGGCTTAGCGGTGTGCCAGGCCAGCTGGGA 989
 QY 482 CTCGGGAGCTCCAGGCTTGCTGGGTACAGGATGCCAGGCCCAAGGGCCCCCCCCCG 541
 Db 990 CAAAGAGGAGCCCTGGAGACCAAGGTGAGCGGGCCCGCAGGGGCTTCCTTGGATTCTGT 1049
 QY 542 GCCCTCTGGCCCATCAGGAGCGGTGTGCCCC 574
 Db 1050 GTCCCCCTGGAAAGAGGGAGAGCAGGGGCTC 1082

RESULT 14
 US-09-964-824A-255
 ; Sequence 255, Application US/09964824A
 ; Patent No. US20020102531A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Horrigan, Stephen
 ; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
 ; TITLE OF INVENTION: Sets
 ; FILE REFERENCE: 689290-73

```
; CURRENT APPLICATION NUMBER: US/09/964,824A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 255
; LENGTH: 2542
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964A-824A-255

Query Match          10.3%; Score 106; DB 10; Length 2542;
Best Local Similarity 52.1%; Pred. No. 5.1e-19;
Matches 236; Conservative 0; Mismatches 217; Indels 0; Gaps

Qy 122 TACAAGGCTCCACCGGGCCCCAGGGGTCCAAGAAGGTGCACAGAGGATCCCAGGGACCCCTG 181
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 630 TCCTTGACCACCGGGTCCCGCAGGCATCAGGGCTACCCAGGCATGSCAGSGGCCAAGG 689

Qy 182 GCCCACTGGCAACAAGGACAGAAAGAGAGAAAGGGAGCGCTGGACCACCTGGCCCTG 241
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 690 GAGAGACGGGCCCTCATGGATATAAAGGCATGGTGGGGCTATCGGTGCCACTGGGCC 749

Qy 242 CGGGTCGAGAGAGGCCCAATTGGACCAAGCTGGTCCCGCGGAGAGCGTGGCGGCAAGGAT 301
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 750 CGGGTAGGAAGGTCTCTAGGGACCGCCAGGCGGAGTGGGGAGAAGGGTGACGAGGGCA 809

Qy 302 CTAAGAGCTCCACGGGCCCCAAAGGCTCCCGTGGTTCCCTCGGAAGCCCGGCCCTCAGG 361
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 810 GCCCAGGTATCTGTGACCCAGGGGATCACAGGCCCGAAAGGACCAACGGSCCCCCCAG 869

Qy 362 GCCCAGTGGGGACCCAGGCCCCCGGGGCCACACAGGCAAGAGAGGACTCCCGGGCCCTG 421
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 870 GCATCAACGGCAAGGATGGGAGCCCGCAGGCACGCTGGCATGAAGGGCAGTGCAGSACAG 929

Qy 422 AGGGCCCTCTGGCTTCCAGGGAGCTTCACGGGCACCGTTGGGGAGCGCTGGGTGCTGGAC 481
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 930 CGGGACAGCCCGGAATCCAGGCCACCGAGGGCTAGCGGGTGTGCCAGGCCAGCTGGGA 989

Qy 482 CTGGGGAGCTGCCAGGCTTTCCTTGGGGTACCAAGGCATGCCAGGCCCAAGGGCCCCCCC 541
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 990 CAAGAAGGGCCCTGCGAGACCAAGGGTGAAGCGGGGCCCGCAGGGCCCTCTCGATTCTG 1049

Qy 542 GCCCTCTGGCCCCATCAGGAGCGGTGGTCCCC 574
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1050 GTCCCCCTGGGAAAAGGGAGAGCCAGGGGCC 1082


RESULT 15
US-09-954-456-786
; Sequence 786, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agen
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
```

```
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 786
; LENGTH: 5416
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-786

Query Match      10.3%; Score 105.4; DB 10; Length 5416;
Best Local Similarity 51.1%; Pred. No. 1.1e-18;
Matches 247; Conservative 0; Mismatches 236; Indels 0; Gaps 0;

QY 120 ACTACAAGGTCACCGGCCCCAGGGGTCACAGAGGTACAGAGATCCCAGGGACCCCC 179
Db 1039 ACAGCCCGGTGCTCTGTGTGAAGGTGAACCTGGTGCCTCTGTAATAATGGAACCTCC 1098

QY 180 TGGCCCAACAGGACAGACAAAGAGAGAGAGGGGAGCCTGGACCACTGGGCC 239
Db 1099 AGGTCAACAGAGAGCCCTGGTCTTCCTGGTGAGAGAGAGCTGTTGGTGCCCTGGTCC 1158

QY 240 TCGGGGTGAGAGAGCCCAATTGGACAGCTGGTCCCGGAGAGAGCGTGGCGGCAAGG 299
Db 1159 AGCTGGTCCCGTGAAGTGTGAAGTGTGGGTCCCGTAGTCTCTGGTCTCTAATGG 1218

QY 300 ATCTAAAGGCTCCAGGCCCCCAAGAGCTCCGGTGGTTCCCTGGGAAGCCCGGCCCTCA 359
Db 1219 GTCTGCTGGCCCTCCAGGTTTCCAGAGTGCCCTGGTCCCAAGGTTGAATTTGAGCTGT 1278

QY 360 GGGCCCCAGTGGGACCCAGGCCCCCGGCCACCAGGCAAGAGGACTCCCGGGCCC 419
Db 1279 TGGTAACGCTGGTCTACTTGACCCCGCGTCCCGTGGTGAAGTGGGTCTTCCAGGCT 1338

QY 420 TCAGGGCCCTCTGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCTGGGGTGCTTG 479
Db 1339 CTCGGCCCGGTGGACCTCCTGGTAATCCTGGAGCAAGGSCCTTACTGGTGCCAAGGG 1398

QY 480 ACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAGGATGCCAGGCCCCCAAGGGCCCCC 539
Db 1399 TGCTGCTGGCCCTTCCCGCGTGTCTGGGGCTCCCGGCTCCCTGGACCCCGCGGTATTCC 1458

QY 540 CGGCCCTCTGGCCCATCAGGAGCGGTGTGCCCTGGCCCTGCAGATGAGCCCAACCCC 599
Db 1459 TGGCCCTCTGGTGTGCCGTACTACTGTGTGCCAGAGGACTGTTGGTGAGCCTGGTCC 1518

QY 600 GGC 602
Db 1519 AGC 1521
```

Search completed: March 21, 2003, 08:22:40
Job time : 175.339 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:34:06 ; Search time 1346.02 Seconds
(without alignments)
12344.923 Million cell updates/sec

Title: US-09-763-712a-1_copy_670_1695

Perfect score: 1026

Sequence: 1 atgcaacaagattgatgag.....cagtactgtcatctgcatta 1026

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estlin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_othr:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 800.4 | 78.0 | 3305 | 11 BC009162 | BC009162 Mus muscu |
| 2 | 556.4 | 54.2 | 861 | 14 B0713873 | B0713873 AGENCOURT |
| 3 | 525.4 | 51.2 | 668 | 14 B0674807 | B0674807 AGENCOURT |
| 4 | 475.4 | 46.3 | 967 | 9 AL568743 | AL568743 AL568743 |
| 5 | 462.6 | 45.1 | 906 | 13 B1456109 | B1456109 60317265 |
| 6 | 445.6 | 43.4 | 580 | 12 BE910803 | BE910803 601661855 |

| | | | | | |
|----|-------|------|------|-------------|--------------------|
| 7 | 437.2 | 42.6 | 552 | 10 AW958053 | AW958053 EST370123 |
| 8 | 415.8 | 40.5 | 601 | 10 BE290299 | BE290299 601089246 |
| 9 | 403.4 | 39.3 | 957 | 14 B0955927 | B0955927 AGENCOURT |
| 10 | 369.4 | 36.0 | 638 | 10 BB248064 | BB248064 BB248064 |
| 11 | 351.4 | 34.2 | 500 | 14 BM676508 | BM676508 UI-E-E30- |
| 12 | 350.4 | 34.1 | 507 | 14 BM713891 | BM713891 UI-E-E30- |
| 13 | 343.8 | 33.5 | 808 | 14 BQ771366 | BQ771366 UI-N-F10- |
| 14 | 311.3 | 30.3 | 339 | 9 AA304251 | AA304251 EST17006 |
| 15 | 301.4 | 29.4 | 357 | 10 AW379436 | AW379436 CM4-HT024 |
| 16 | 297.2 | 29.0 | 517 | 9 AL543000 | AL543000 AL543000 |
| 17 | 297 | 28.9 | 936 | 14 BQ891432 | BQ891432 AGENCOURT |
| 18 | 296 | 28.8 | 310 | 9 AA361740 | AA361740 EST17069 |
| 19 | 271 | 26.4 | 350 | 9 AA012704 | AA012704 RPU1402CG |
| 20 | 245.8 | 24.0 | 420 | 12 BE929813 | BE929813 RC3-GN004 |
| 21 | 233.8 | 22.8 | 429 | 12 BE929798 | BE929798 RC3-GN004 |
| 22 | 212 | 20.7 | 723 | 14 BQ004176 | BQ004176 UI-H-E10- |
| 23 | 210.2 | 20.5 | 943 | 14 BQ887163 | BQ887163 AGENCOURT |
| 24 | 201 | 19.6 | 655 | 12 BG482931 | BG482931 602502939 |
| 25 | 200.8 | 19.5 | 345 | 9 A1742661 | A1742661 wg44f06.x |
| 26 | 199.6 | 19.5 | 441 | 10 BE749743 | BE749743 BB749743 |
| 27 | 185.2 | 18.1 | 591 | 13 B1445884 | B1445884 dal33d10. |
| 28 | 185 | 18.0 | 277 | 12 BF854320 | BF854320 MR2-EN009 |
| 29 | 183.4 | 17.9 | 234 | 12 BF854188 | BF854188 MR2-EN009 |
| 30 | 183.4 | 17.9 | 261 | 13 B1010880 | B1010880 MR2-EN009 |
| 31 | 180.2 | 17.6 | 280 | 13 B1010883 | B1010883 MR2-EN009 |
| 32 | 178.8 | 17.4 | 286 | 12 BF854324 | BF854324 MR2-EN009 |
| 33 | 172.4 | 16.8 | 334 | 10 AW002141 | AW002141 wt84a12.x |
| 34 | 170.4 | 16.6 | 309 | 14 W76214 | W76214 zq58e09.r1 |
| 35 | 159.8 | 15.6 | 281 | 12 BF854408 | BF854408 MR2-EN009 |
| 36 | 151.2 | 14.7 | 465 | 10 AW240221 | AW240221 up30b07.y |
| 37 | 144.4 | 14.1 | 504 | 14 R74387 | R74387 y157f11.r1 |
| 38 | 142.8 | 13.9 | 288 | 14 W72977 | W72977 zd58e09.s1 |
| 39 | 135.6 | 13.2 | 432 | 9 A1613616 | A1613616 vg05h05.y |
| 40 | 127.2 | 12.4 | 1082 | 14 BM907108 | BM907108 AGENCOURT |
| 41 | 124.4 | 12.1 | 884 | 14 BQ934501 | BQ934501 AGENCOURT |
| 42 | 124.2 | 12.1 | 531 | 10 AW603727 | AW603727 AV603727 |
| 43 | 123.8 | 12.1 | 608 | 10 AW918393 | AW918393 EST349697 |
| 44 | 122.4 | 11.9 | 960 | 14 BQ920508 | BQ920508 AGENCOURT |
| 45 | 118.8 | 11.6 | 824 | 13 B1828870 | B1828870 603075026 |

ALIGNMENTS

| | | | | | | |
|------------|---|--|---------|------|--------|-----------------|
| RESULT 1 | BC009162 | Mus musculus, Similar to scavenger receptor with C-type lectin, clone IMAGE:2811487, mRNA. | 3305 bp | mRNA | linear | HTC 07-AUG-2002 |
| LOCUS | BC009162 | HTC | | | | |
| DEFINITION | BC009162 | HTC | | | | |
| ACCESSION | BC009162.1 | GI:14714370 | | | | |
| VERSION | BC009162 | | | | | |
| KEYWORDS | HTC | | | | | |
| SOURCE | house mouse. | | | | | |
| ORGANISM | Mus musculus | | | | | |
| REFERENCE | 1 (bases 1 to 3305) | | | | | |
| AUTHORS | Strausberg, R. | | | | | |
| TITLE | Direct Submission | | | | | |
| JOURNAL | Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA | | | | | |
| REMARK | NIH-MGC Project URL: http://mgc.nci.nih.gov | | | | | |
| COMMENT | Contact: MGC help desk Email: cgabbs-remail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ | | | | | |

Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 5 Row: h Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis

This clone has the following problem: incomplete processing.

FEATURES

source

Location/Qualifiers

1..3305

/organism="Mus musculus"

/db_xref="taxon:10090"

/map="C57BL/6J"

/clone="IMAGE:2811487"

/tissue_type="mammary tumor. WAP-rtGf alpha model. 7 months

old, gross tissue."

/clone_lib="NCI CGAP_Mam5"

/lab_host="DH10B"

/note="Vector: PCMV-SPORT6"

BASE COUNT 1031 a 757 c 760 g 757 t

ORIGIN

Query Match 78.0%; Score 800.4; DB 11; Length 3305;
Best Local Similarity 86.3%; Pred. No. 2.4e-179;
Matches 885; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
QY 1 ATGCAACAAGATTGATGAGGTGCGAGGTGACACTGAAGTAGCCCAACTTATCAGTGATT 60
DB 1296 ATGCAGCAAGACATGATGAGGTCAAAGTTAGACACTGAAGTAGCCCAACTTATCAGTGGTT 1355
QY 61 ATGGAAGAATGAAGTAGTACAGTCCAGCATGCTCAGCTCATCAAGAAATTTTACAATA 120
DB 1356 ATGGAAGATGAAGTAGTGTGACTCCAAAGCACGGTCACTCATCAAGAACTTTTACCAATT 1415
QY 121 CTACAGGTGCCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGGATCCCAAGGACCCCT 180
DB 1416 CTACAGGTGCTCTCTGGCCCCAGAGGTCCAAAGGTGACAGAGGATCTCAGGACCACT 1475
QY 181 GCGCCCACTGGCAACAGAGGACAGAAAGAGAGAGGGGAGCCCTGGACCACTTGGCCCT 240
DB 1476 GTTCCAACCTGGCAACAGAGGACAGAAAGAGAGAGGGAGAGCCCTGGTCCACCTGGCCCT 1535
QY 241 GCGGGTGAGAGAGGCCAATTGGACACGCTGTGCCCCGGAGAGCGTGGCGGCAAGGA 300
DB 1536 GCGGGTGAGAGAGGACACATTTGGACACGCTGGCCCTCTCTGGAGAGCGTGGCAAGGA 1595
QY 301 TCTAAGGTCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCCGTGGGAAGCCCGGCTCAG 360
DB 1596 TCCAAGGCTCAGAGGTCCCAAGAGATCTCGTGGTCCCAAGGAGAGCGTGGCCCTCAA 1655
QY 361 GCGCCCACTGGGACCCAGGCCCCCGGGCCCAACAGGCAAGAGAGGACTCCCGGCGCT 420
DB 1656 GGACCTAGTGGGGAGCCAGGACCAAGGATCCACAGGCAAGAGTGGACTCCCTGGCCCT 1715
QY 421 CAGGCGCTCTCTGGCTTCCAGGACTTACAGGCACCGTGGGGAGCGCTGGGCTGCTGA 480
DB 1716 CAGGCGCTCTCTGGCTTCCAGGACTTACAGGCACTGTGGGTGGAGCTGGAGTACTGGA 1775
QY 481 CCGGGGACTCCAGGCTTGGCTGGGTACCAGGCATGCCAGGCCCCCAAGGCGCCCCC 540
DB 1776 CCGGGGCTTCCAGGCTTGGCAGGGGTGCCAGGCATGCCCTGGGCTTAAGGACCACT 1835
QY 541 GCGCCCTCTGGCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAAGAGGACCCCG 600
DB 1836 GCGCCCTCAGGCGCCCTCAGGAGCAATGGAGCAATTGGCTCTGCAGAAATGAACCAACCCA 1895
QY 601 GCACGGAGGACAATGGCTGGCCGCTCACTGGAAGACTTCACAGACAAATGCTACTAT 660
DB 1896 GCATCAGAGGTACAGGATGTCCACTCACTGGAAGACTTCACAGATAAATGCTACTAT 1955
```

```
QY 661 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTCA 720
DB 1956 TTTTCATTGGAAGAAAGAAATTTTGAAGATGCTAAAGCTTTTCTGTGAAGACAAGTCTTCC 2015
QY 721 CATCTGTTTTCATTAACACTAGAGAGGAACAGCAATGATTAATAAAGACAGATGGTAGGG 780
DB 2016 CATCTGTTTTCATTAACACTCAAGAGAAAGAACAGCAATGATTAATAAAGACATACCGTGGG 2075
QY 781 AGAGAGAGCACTGGATCGGCTCACAGACTCAGAGCGTGAAAAATGAATGGAAGTGGCTG 840
DB 2076 AGAGAAAGCCATGGATCGGCTCACAGACTCAGAACAGGAAGCAATGGAAGTGGCTA 2135
QY 841 GATGGACATCTCCAGACTTACAAAATTTGAAAGCTGGACAGCCGATTAACCTGGGTCTAT 900
DB 2136 GACGGTCACTGTTGATTACAAAACCTGAAAAGCTGGACACACAGATAAAGTGGGCACT 2195
QY 901 GGCATGGCCAGGAGAGAGACTGTGCTGGTTGATTATGCTGGGAGTGGAGACGATTTC 960
DB 2196 GGCATGGCCAGGAGAGAGACTGTGCTGGTTGATTATGCTGGGAGTGGAGTGGATGACTTC 2255
QY 961 CAATGTGAAGACGTCAATAACTTCAATTCGAAAAAGACAGGAGACAGTGTGCTCATCT 1020
DB 2256 CAGTGTGATGAATCAATAACTTCAATTCGAAAAAGAGGAGGAGGACGATCATCATCC 2315
QY 1021 GCATTA 1026
DB 2316 ATATTA 2321
```

```
RESULT 2
BQ713873
LOCUS BQ713873 861 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8482828 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6306240
5', mRNA sequence.
ACCESSION BQ713873
VERSION BQ713873.1 GI:21852772
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 861)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: ResGen, Invitrogen Corp
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13721 row: n column: 01
High quality sequence stop: 598.
```

FEATURES

source

1..861

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:6306240"

/clone_lib="NIH_MGC_129"

/lab_host="DH10B (phage-resistant)"

/note="Organ: olfactory epithelium; Vector:

PCMV-SPORT6.1.ccd; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dr. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH_MGC Library."

BASE COUNT 221 a 253 c 239 g 147 t 1 others

ORIGIN

Query Match 54.2%; Score 556.4; DB 14; Length 861;

| | |
|--|---|
| Best Local Similarity 82.8%; Pred. No. 1.4e-121; Matches 647; Conservative 0; Mismatches 132; Indels 2; Gaps 1; | |
| Qy 14 | TGATGAGGTCTAGACACTGAAGTAGCCAACTTATCAGTGATATGGAAGAAATGA 73 |
| Db | |
| Qy 14 | TGATGAGGTCAAGTTAGACACTGAAGTGCCCACTTATCAGTGATGGAAGAGATGA 73 |
| Db | |
| Qy 74 | AGCTAGTAGATCCAAAGCATGGTCAGCTCATCAAGATTTTACAAATCTACAAGGTCCAC 133 |
| Db | |
| Qy 74 | AACGTGTGACTCCAAAGCACGGTCAGCTCATCAAGATTTTACCAATCTACAAGGTCCAC 133 |
| Db | |
| Qy 134 | CGGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCAGGAGACCCCTGGCCCCCACTGCCA 193 |
| Db | |
| Qy 134 | CTGGCCCCAGAGGTCCAAAGAGGTGACAGAGATCTCAGGAGACCACTGGTCCAACTGCCA 193 |
| Db | |
| Qy 194 | ACAAGGGACAGAAAGGAGAGAGGGGAGCTTGGACCACTTGGCCCTCGGGGTGAGAGAG 253 |
| Db | |
| Qy 194 | ACAAGGGACAGAAAGGAGAGAGGGAGAGCTTGGTCCACCTGGCCCTCGGGGTGAGAGG 253 |
| Db | |
| Qy 254 | GCCCAATTGGACAGCTGCTCCCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCC 313 |
| Db | |
| Qy 254 | GCACAAATTGGACAGCTGCTCCCGGAGAGCGTGGCGGCAAGAGATCTAAAGGCTCCAC 313 |
| Db | |
| Qy 314 | AGGGCCCCAAGAGCTCCCGTGGTCCCTGGGAGAGCCCGGCCCTCAGGGCCCCAGTGGGG 373 |
| Db | |
| Qy 314 | AGGGTCCCAAGAGATCTCGTGGGTCCCGGAGGAGAGCTTGGCCCTCAAGGACCTAGTGGGG 373 |
| Db | |
| Qy 374 | ACCAGGGCCCCCGGCCCCACAGCAAGAGGAGTCCCGGCCCTCAGGGCCCCCTCTCG 433 |
| Db | |
| Qy 374 | ACCAGGACCAAGGATCCACAGCAAGAGATGAGTCCCTGGCCCTCAGGGCCCCCTCTCG 433 |
| Db | |
| Qy 434 | GCTTCCAGGGATCTCAGGGACCTTGGGGAGCCCTGGGGAGCCCTGGAGCTCGGGGACTGC 493 |
| Db | |
| Qy 434 | GCTTCCAGGGACTACAGGGACTGTGGGTGAGCCTGGAGTACCTGGACCTCGGGGGTTGC 493 |
| Db | |
| Qy 494 | CAGGCTTGCCTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGCCCTCTCTGGCC 553 |
| Db | |
| Qy 494 | CAGGCTTGCCTGGGTGAGGAGTGCCTGGGGCTTAAGGGACCACTTGGGCCCTCAGGGCC 553 |
| Db | |
| Qy 554 | CATCAGGAGCGGTGGTCCCTGGCCCTGCAGAAATGAGCCCAAGCCCGGACCGGAGACA 613 |
| Db | |
| Qy 554 | CCTCAGGAGCAATGGAGCCATGGCTCTGCAAAATGAAACCAACCCCAACATCAGAGTCA 613 |
| Db | |
| Qy 614 | ATGCTGCCCCCTCTACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGA 673 |
| Db | |
| Qy 614 | ACGGATGTCGCCCTCTACTGGAAGAACTTCACAGATAAATGCTACTATTTTTCCTTGGCAA 673 |
| Db | |
| Qy 674 | AAGAAATTTTGGAGTCAAGC--TTTCTGTGAAGACAAGTCTTCACATCTTGTTT 731 |
| Db | |
| Qy 674 | AAAAAATTTTTCGAAATGCTTACCTTTTCTGCGAACAAAAATCTCCCCCACTCCGGGTTT 733 |
| Db | |
| Qy 732 | CATAAACAATAGAGGAACAGCAATGATATAAAAAACAGATGTTAGGGAGAGAGAGCCA 791 |
| Db | |
| Qy 734 | CATACCCCCCAAGAAAAACAGCCATGGGATAAAAAACCTTACCTTCGGGGACCAAGAAC 793 |
| Db | |
| Qy 792 | C 792 |
| Db | |
| Qy 794 | C 794 |
| Db | |
| RESULT 3 | |
| BQ674807 | |
| LOCUS | |
| DEFINITION BQ674807 668 bp mRNA linear EST 15-JUL-2002 | |
| AGENCY NCBI | |
| 5', mRNA sequence. | |
| ACCESSION BQ674807 | |
| VERSION BQ674807.1 | |
| KEYWORDS EST. | |
| SOURCE human. | |
| ORGANISM Homo sapiens | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| REFERENCE 1 (bases 1 to 668) | |

| | |
|---|---|
| AUTHORS NIH-MGC http://mgc.nci.nih.gov/. TITLE National Institutes of Health, Mammalian Gene Collection (MGC) JOURNAL Unpublished (1999) COMMENT Contact: Robert Strausberg, Ph.D. Email: cgaaps-r@mail.nih.gov Tissue Procurement: ATCC | |
| cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LICM2377 row: c column: 09 High quality sequence stop: 656. | |
| FEATURES Location/Qualifiers 1. 668 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:6212672" /tissue_type="epidermoid carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: salivary gland; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library." | |
| BASE COUNT | 216 a 148 c 167 g 137 t |
| ORIGIN | |
| Query Match 51.2%; Score 525.4; DB 14; Length 668; Best Local Similarity 99.8%; Pred. No. 3e-114; Matches 526; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | |
| Qy 500 | TGCCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCTCTGGCCCATCAG 559 |
| Db | 1 TGCCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCTCTGGCCCATCAG 60 |
| Qy 560 | GAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCCAACCCCGGACCGGAGGACATGGCT 619 |
| Db | 61 GAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCCAACCCCGGACCGGAGGACATAGCT 120 |
| Qy 620 | GCCGCGCTCACTGGAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAGAAA 679 |
| Db | 121 GCCGCGCTCACTGGAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAGAAA 180 |
| Qy 680 | TTTTTGAGGATGCAAGGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTCATTAACA 739 |
| Db | 181 TTTTGGAGATGCAAGGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTCATTAACA 240 |
| Qy 740 | CTAGAGAGAACAGCAATGGATATAAAACAGATGGTAGGAGAGAGACCACTGGATCG 799 |
| Db | 241 CTAGAGAGAACAGCAATGGATATAAAACAGATGGTAGGAGAGAGACCACTGGATCG 300 |
| Qy 800 | GCCTCACAGCTCAGAGCGTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 859 |
| Db | 301 GCCTCACAGCTCAGAGCGTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 360 |
| Qy 860 | ACAAAAATTTGAAAGCTGGACAGCGCGGATAACTGGGGTCATGGCCATGGCCAGGAGAG 919 |
| Db | 361 ACAAAAATTTGAAAGCTGGACAGCGCGGATAACTGGGGTCATGGCCATGGCCAGGAGAG 420 |
| Qy 920 | ACTGTGCTGGGTTGATTTATGCTGGGCACTGGAACGATTTTCCAAATGTGAAGACGTCATA 979 |
| Db | 421 ACTGTGCTGGGTTGATTTATGCTGGGCACTGGAACGATTTTCCAAATGTGAAGACGTCATA 480 |
| Qy 980 | ACTTCATTTGGGAAAAACAGAGGAGACAGTACTGTCATCTGCATTA 1026 |
| Db | 481 ACTTCATTTGGGAAAAACAGAGGAGACAGTACTGTCATCTGCATTA 527 |

```
RESULT 4
AL568743/c
LOCUS      AL568743      967 bp      mRNA      linear      EST 16-FEB-2001
DEFINITION AL568743 LTI_FL002_PL1 Homo sapiens cDNA clone CS0DE005YH04 3 prime
            , mRNA sequence.
ACCESSION  AL568743
VERSION     AL568743.1 GI:12923387
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 967)
AUTHORS     Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
CONTACT     Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES             Location/Qualifiers
     source           1..967
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DE005YH04"
                     /clone_lib="LTI_FL002_PL1"
                     /lab_host="DH10B"
                     /note="organ: placenta; Vector: pCMVSPORT 6; 1st strand
                     cDNA was primed with a NotI-oligo(dT) primer. Five prime
                     end enriched, double-stranded cDNA was digested with Not I
                     and cloned into the Not I and Eco RV sites of the
                     pCMVSPORT 6 vector. Library was constructed by Life
                     Technologies. Contact : Feng Liang Life Technologies, a
                     division of Invitrogen 9800 Medical Center Drive Rockville
                     Maryland 20850, USA Fax : (1) 301 610 8371 Email :
                     filiang@life.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      222 a      221 c      217 g      290 t      17 others
ORIGIN
Query Match      46.3%; Score 475.4; DB 9; Length 967;
Best Local Similarity 96.7%; Pred. No. 2.6e-102;
Matches 504; Conservative 3; Mismatches 9; Indels 5; Gaps 2;

QY 506 GGGTACAGGCGATGCCAGGCCCAAGGCCCGCCCGCCCTCTGCCCCATCAGGAGCGG 565
      |||||
DB 898 GGGTACAGGCGATGCCAGGCCCAAGGCCCGCCCGCCCTCTCT---GCCATCAGGAGCGG 843
      |||||
QY 566 TGGTGGCCCTGGCCCTGCAGATGAGCCAAACCCCGGCACCGGAGGACAAATGGCTGCCCGC 625
      |||||
DB 842 TTGTGCCCTTGGCCCTGCAGATGAGCCAAACCCCGGCACCGGAGGACAAATGGCTGCCCGC 783
      |||||
QY 626 CTCACCTGGAAGAACTTCACAGACAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTG 685
      |||||
DB 782 CTCACCTGGAAGAACTTCACAGACAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTG 723
      |||||
QY 686 AGGATCAAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTTTTCATAAACAACACTAGAG 745
      |||||
DB 722 AGGATCAAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTTTTCATAAACAACACTAGAG 663
      |||||
QY 746 AGAACAGCAATGGATAAATAAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCCTCA 805
      |||||
DB 662 AGAACAGCAATGGATAAATAAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCCTCA 603
      |||||
QY 806 CAGACTCAGAGCGTGAATGAATGAATGAGTGGTGGTGGGACATCTCCAGACTACAAAA 865
      |||||
DB 602 CAGACTCAGAGCGTGAATGAATGAATGAGTGGTGGTGGGACATCTCCAGACTACAAAA 543
      |||||
QY 866 ATTGGAAGCTGGACAGCGCGGATAACTGGGGTCATGGCCATGGCCAGGAGAGAGACTGTG 925
      |||||
DB 542 ATTGGAAGCTGGACAGCGCGGATAACTGGGGTCAT-GCCATGGCCAGGAGAGAGACTGTG 484
      |||||
```

```
QY 926 CTGGGTTGATTTATCTGGGCAGTGGACGAGATTCCAATGTGAAGACGTCAATAACTTCA 985
      |||||
DB 483 CTGGGTTGATTTATCTGGGCAGTGGACGAGATTCCAATGTGAAGACGTCAATAACTTCA 424
      |||||
QY 986 TTTGGAAAAAGACAGAGGAGACAGTACTCTCATCTGCATTA 1026
      |||||
DB 423 TTTGGAAAAAGACAGAGGAGACAGTACTCTCATCTGCATTA 383
      |||||

RESULT 5
BI456109      906 bp      mRNA      linear      EST 21-AUG-2001
LOCUS      603172765F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5251888 5',
DEFINITION mRNA sequence.
ACCESSION  BI456109
VERSION     BI456109.1 GI:15246765
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 906)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11636 row: b column: 17
            High quality sequence stop: 730.

FEATURES             Location/Qualifiers
     source           1..906
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5251888"
                     /clone_lib="NCI_CGAP_Mam5"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="7 months"
                     /lab_host="DH10B"
                     /note="organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                     Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys,
                     NIH"
BASE COUNT      247 a      228 c      256 g      175 t
ORIGIN
Query Match      45.1%; Score 462.6; DB 13; Length 906;
Best Local Similarity 81.7%; Pred. No. 2.8e-99;
Matches 597; Conservative 0; Mismatches 124; Indels 10; Gaps 5;

QY 1 ATGCAACAAGATTGATGAGTTCGAGGTTAGACACTGAAGTAGCCCAACTATTCAGTGATT 60
      |||||
DB 123 ATGCAGCAAGACATGATGAGTCAAGATTAGACACTGAAGTAGCCCAACTATTCAGTGATT 182
      |||||
QY 61 ATGGAAGAAATGAAGCTAGTAGCTCCAAAGCATGGTTCAGCTCATCAAGAATTTTACAATA 120
      |||||
DB 183 ATGGAAGAGATGAATGGTTGACTTCCAAAGCAGCGTTCAGCTCATCAAGAATTTTACCATT 242
      |||||
QY 121 CTACAGGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCCGAGGACCCCTT 180
      |||||
DB 243 CTACAGGTCTCCCTGGCCCCAGAGGTCCAAAAGGTGACAGAGGATCTCAGGACCACT 302
      |||||
QY 181 GGCCCCAACTGGCAACAAGGGACAGAAAAGGAGAGAGGGGGAGCGCTGGACCACTGGCCCT 240
      |||||
DB 303 GTCCCACTGGCACACAGGGACAGAAAGGAGAGAGAGAGAGCGCTGGTCCACCTGGCCCT 362
      |||||
```

```

241 CGGGTGAGAGGCCCCAATTGGACCACTGGTCCCTCCCGGAGAGCGTGGCGGCAAGGA 300
    ||||| ||| ||||| ||||| ||| ||||| ||||| ||| ||||| ||||| |||||
Db  CGGGTGAGAGGCGCACAAATGGACCACTGGCGCTCTCTCGAGAGCGTGGCGCAAGGA 422
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  TCTAAGGCTCCAGGCGCCCAAGGCTCCCGTGGTTCCTCCCTGGGAGCGCGGCGCTCAG 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  TCCAAAGGCTCACAGGGTCCCAAGGATCTCGTGGGTCCCGAGTGAAGCTTGGCCCTCA 482
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  GGGCCAGTGGGAGCCACAGGCGCCCGGCGCCACAGCAAGAGGAGTCCCGGCGCT 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  GGACCTAGTGGGAGCCACAGGACCAAGTCCACAGCAAGGATGGACTCCCTGGCGCT 542
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  CAGGCGCTCTCGCTTCCAGGCACTTACAGGCACTGAGGCGCTGGAGTACCTTGA 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  CAGGCGCTCTCGCTTCCAGGCACTTACAGGCACTGAGGCGCTGGAGTACCTTGA 602
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  CCTGGGAGCTGCCAGGCTTCCCTGGGTACAGGATGCCAGGCGCCCGGCGCGCC 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  CCTGGGAGCTGCCAGGCTTCCCTGGGTACAGGATGCCAGGCGCCCGGCGCGCC 662
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  GGGCCCTCTCGGCGCATCAGGAGCGGTGGTGGCGCTGCGGCTGCGAGATGAGCCACCG 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  GG-CCTCAGGCGCTCAGGAGC-ATGGAGCCATTTGCTGCAGAAATGAACCAAGCCCA 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  GCACCGGAGGACAATGCG--TGCCCGCGCTCACTGGAAGAACTTCACA---GACAAATGCT 655
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  GCATCAGAGGTTCAACGGATGTCCTGCTCACTGGAAGAACTTCACGGGATTAATGCT 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  ACTATTTTCACTTG--AGAAAGAAATTTTGAGGATGCAAGCTTTCTGTGAAGACA 712
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  ACTATTTTTCATTCACAAAGGAAATTTTGAAGATTTGTAAGCGTTCTGTGAGAAA 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy  713 AGTCTTCACAT 723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  841 AATTTTCCAA 851
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
BE910803 580 bp mRNA linear EST 29-SEP-2000
LOCUS 601661855F1 NCI_CGAP_Maml Mus musculus cdna clone IMAGE:3962292 5',
DEFINITION mRNA sequence.
ACCESSION BE910803
VERSION BE910803.1 GI:10407765
KEYWORDS EST, house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 580)
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9128 row: m column: 13
High quality sequence stop: 580.
Location/Qualifiers
1. 580
/organism="Mus musculus"
/strain="FVB/N"
/db.xref="taxon:10090"
/clone="IMAGE:3962292"
/clone_lib="NCI_CGAP_Maml"
/tissue_type="tumor, biopsy sample"

```

```

/dev_stage="10 months, virgin"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site:1; SalI;
Site:2; NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 159 a 153 c 161 g 107 t
ORIGIN

Query Match 43.4%; Score 445.6; DB 12; Length 580;
Best Local Similarity 85.5%; Pred. No. 2.6e-95;
Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 290 CGCGCAAGAGATCTAAAGGCTCCAGGCGCCCAAGGCTCCCGTGGTTCCTCCCTGGGAAGC 349
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 CGAGCAAGAGATCCAAAGGCTCACAGGGTCCCAAGGATCTCGTGGTCCCCAGGGAAGC 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 350 CGGCGCTCAGGCGCCCAAGTGGGACCCAGGCGCCCGGCGCCACAGGCAAGAGGAC 409
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 CTGGCCCTCAAGGACCTAGTGGGACCCAGGACCCAGGTCACAGGCAAGGATGGAC 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 410 TCCCGGCGCTCAGGCGCTCTCTGGCTTCCAGGGACTTCAGGGACCGTTGGGGAGCGCTG 469
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 TCCCTGGCGCTCAGGCGCTCTCTGGCTTCCAGGGACTTCAGGGACCGTTGGGTGAGCGCTG 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 470 GGGTGCCTGAGACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCGAGGCATGCCAGGCCCA 529
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 GAGTACCTGGACCTCGGGGTTGCCAGGCTTGGCAGGGGTGCCAGGCATGCTGGGCGCTA 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 530 AGGGCGCGCGCGCGCTCTCTGGCGCATCAGGCGGGTGGTGGCGCTGGCCCTGCAAGATG 589
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 AGGGACCACTTGGCGCTCCAGGCGCTCCAGGCGCTCCAGGCAAGTGGAGCCATTTGCTGCAAGATG 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 590 AGCCAAACCCGCGACCGGAGGACATGCTGCCGCTCAGTGGAGAACTTCACAGACA 649
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 AACCAACCCGAGCATCAGAGGTCAACGGATGTCCGCTCAGTGGAGAACTTCACAGATA 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 650 AATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGTGCAGAAAGCTTTTCTGTGAAG 709
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 AATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGTGCAGAAAGCTTTTCTGTGAAG 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 710 ACAAGTCTTCACATCTTTTTCATAAACAACATAGAGAGGACAGCAATGGATGATAAAGAAC 769
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 ACAATCTTCCCATCTCGTTTTCATAAACAACATAGAGAGGACAGCAATGGATGATAAAGAAC 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 770 AGATGGTAGGAGAGAGCCACCTGGATCGGCTCAGAGCTCAGAGCGTGAATGAATGAAT 829
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 ATACCGTGGGAGAGAGAGCCATTGGATCGGCTCAGAGCTCAGAGCGTGAATGAATGAAT 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 830 GGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTG 869
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 GGAAGTGGCTAGAGGGTCACTGTTGATTACAAAAACTG 580
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
AW958053 552 bp mRNA linear EST 01-JUN-2000
LOCUS EST370123 MAGE resequences, MAGE Homo sapiens cdna, mRNA sequence.
DEFINITION AW958053
ACCESSION AW958053
VERSION AW958053.1 GI:8147736
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 552)
TITLE Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspar,R., Gay,C., Holt
JOURNAL ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
COMMENT Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
Contact: John Quackenbush

```

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 115
Seq primer: Reverse.

FEATURES

Location/Qualifiers

1..552

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="MAGE resequences, MAGE"

/note="Vector: pBluescriptSKm"

176 a 112 c 143 g 121 t

BASE COUNT

Query Match 42.6%; Score 437.2; DB 10; Length 552;
Best Local Similarity 99.1%; Pred. No. 2.5e-93;
Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

1;

QY 574 CTGGCCCTGCAGAAATGAGCAACCCGCGGAGGAGCAATGGCTGCCCGCTCACTGG 633

Db 1 CTGGCCCTGCAGAAATGAGCAACCCGCGGAGGAGCAATGGCTGCCCGCTCACTGG 60

QY 634 AAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGAGGATGCA 693

Db 61 AAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGAGGATGCA 120

QY 694 AAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTGTTCATTAACACATAGAGAGGAACAG 753

Db 121 AAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTGTTCATTAACACATAGAGAGGAACAG 180

QY 754 CAATGGATAAAAAACAGATGTTAGGAGAGAGCCACTGGATCGGCTCAGACTCA 813

Db 181 CAATGGATAAAAAACAGATGTTAGGAGAGAGCCACTGGATCGGCTCAGACTCA 240

QY 814 GAGCGTGAATGAATGAGTGGCTGATGGAGATCTCCAGACTACAAAAATTGGAAA 873

Db 241 GAGCGTGAATGAATGAGTGGCTGATGGAGATCTCCAGACTACAAAAATTGGAAA 300

QY 874 GCTGACAGCCGATTAACCTGGGGTTCATGGCCATGGCCAGGAGAGACTGTCTGGGTG 933

Db 301 GCTGACAGCCGATTAACCTGGGGTTCATGGCCATGGCCAGGAGAGACTGTCTGGGTG 360

QY 934 ATTATGCTGGCAGTGAACGATTTCCTCAATGTGAAGAGCTCAATAAC-TTCATTGGGA 992

Db 361 ATTATGCTGGCAGTGAACGATTTCCTCAATGTGAAGAGCTCAATAAC-TTCATTGGGA 420

QY 993 AAAACACAGGGAGACAGTACTGTCTGTCATTA 1026

Db 421 AAAACACAGGGAGACAGTACTGTCTGTCATTA 454

RESULT 8

BE290299

LOCUS

DEFINITION 601089246F1 NCI_CGAP_Mam5 Mus musculus cdna clone IMAGE:3484163 5',

mRNA sequence.

ACCESSION BE290299

VERSION BE290299.1

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 601)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cdna Library Preparation: Life Technologies, Inc.

cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM8516 row: k column: 12
High quality sequence stop: 551.

Location/Qualifiers

1..601

source

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_lib="IMAGE:3484163"

/tissue_type="tumor, gross tissue"

/dev_stage="7 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigators

providing samples: Lothar Hennighausen/Robin Humphreys,

NIH"

BASE COUNT 156 a 161 c 179 g 105 t

ORIGIN

Query Match 40.5%; Score 415.8; DB 10; Length 601;

Best Local Similarity 84.8%; Pred. No. 3.2e-88;

Matches 501; Conservative 0; Mismatches 87; Indels 3; Gaps 3;

QY 196 AAGGACAGAAAGGAGAGAGGGGAGCCTGGACACACCTGGCCCTCGGGGTGAGAGAGGC 255

Db 9 AAGTGCACAGAAAGGAGAGAGGGGAGCCTGGTCCACCTGGCCCTCGGGGTGAGAGAGGC 68

QY 256 CCAATTGGACACAGCTGTCCTCCCGGAGAGCTGGCGGCAAGAGATCTAAAGGCTCCAG 315

Db 69 ACAATTGGACACAGCTGTCCTCCCGGAGAGCTGGCGGCAAGAGATCTAAAGGCTCCAG 128

QY 316 GGGCCCAAGAGCTCCCGTGGTTCCTCTGGG - AAGCCCGGCGCTCAGGGCCCGCCAGGGGA 374

Db 129 GGTCCCAAGAGATCTCGTGGTCCCGGAGGCTGAAGCTGGCCCTCAAGGACCTAGTGGGA 188

QY 375 CCCAGGCCCCCGGCGCCACAGGCAAGAGAGGAGCTCCCGGCGCTCAGGGCCCTCTGG 434

Db 189 CCCAGGACCAACAGTCCACAGGCAAGAGATGGACTCCCTGGCCCTCAGGGCCCTCTGG 248

QY 435 CTTCAGGAGACTTCAGGGCACCGTTGGGAGCTGGGGTGCCTGGACCTCGGGGACTGCC 494

Db 249 CTTCAGGAGACTTCAGGGCACCGTTGGGAGCTGGGGTGCCTGGACCTCGGGGACTGCC 308

QY 495 AGGCTTGGCTGGGGTACCAGGCAATGCCAGGCCCCCAAGGGCCCCCGGCGCTCTGGGCC 554

Db 309 AGGCTTGGC - AGGGTCCAGGCAATGCCAGGCCCCCAAGGGCCCCCGGCGCTCTGGGCC 367

QY 555 ATCAGGAGGGTGGTGGCCCTGGCGCTTCGAGAAATGAGCAACCCCGGCGCAGGAGAGCA 614

Db 368 CTCAGGAGCAATGGAGCAATGGCTCTGCAGAAATGAACCAACCCCGGCGCAGGAGTCAA 427

QY 615 TGGCTGCGCCCTCACTCGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAA 674

Db 428 CGATGTCGCGCTCACTCGAAGAACTTCACAGATTAATGCTACTATTTTTCAGTTGAGAA 487

QY 675 AGAAATTTTGGAGTGCAGAAAGCTTTTCTGTGAAGCAAGTCTTCACATCTTTTTCAT 734

Db 488 AGAAA-TTTTGAAGATGCTAAGCTTTTCTGTGAAGCAAGTCTTCACATCTTTTTCAT 546

QY 735 AAACACTAGAGAGGAACAGCAATGGATTAATAAACAGATGGTAGGAGAGA 785

Db 547 AAACCTCAAGAGAGAAACAGCAATGGTAAAGAAACATACGTGGGGGAGAGA 597

RESULT 9

BO955927

LOCUS

BO955927

957 bp

mRNA

linear

EST 21-AUG-2002

```

DEFINITION  AGENCOURT_8763247 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6313500
5', mRNA sequence.
ACCESSION   B0955927
VERSION     B0955927.1 GI:22371405
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: ResGen, Invitrogen Corp
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM13740 row: 1 column: 13
High quality sequence stop: 665.
FEATURES             source
    source
        1..957
            Location/Qualifiers
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="IMAGE:6313500"
                /clone_11b="NIH_MGC_129"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: (olfactory epithelium); Vector:
                pCMV-SPORT6.1.cdb; Site.1: EcoRV; Site.2: NotI; Cloned
                unidirectionally. Primer: Oligo dT. Average insert size
                2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
                is a NIH_MGC Library."
BASE COUNT  260 a 269 c 258 g 169 t      1 others
ORIGIN
Query Match      39.3%; Score 403.4; DB 14; Length 957;
Best Local Similarity 86.2%; Pred. No. 3.3e-85;
Matches 481; Conservative 0; Mismatches 72; Indels 5; Gaps 3;

Oy 1 ATGCAACAAGATTTGATGAGTTCGAGGTAGACACTGAAGTAGCCAACTTATCATGTTGAT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 ATGCAGCAAGACATGATCAGGTCAAAGTTAGACACTGAAGTGGCCAACTTATCATGTTG 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 61 ATGGAAGAAATGACGTAGTACTCAAGCATGTCAGTCTCATCAAGAATTTACAATA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 413 ATGGAAGAGATGAACACTGGTTGACTCCCAAGCAGCGTCACTCATCAAGAATTTACCAT 472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 121 CTACAAGTCTCCACGGGCCCCAGGGGTCCTCAAGAGTGTACAGAGATCCAGGACCCCT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 473 CTACAGTCTCTCTGGCCCCCAGAGTCCAAAGGTGACAGAGATCTCAGGGACCATC 532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 181 GCGCCAACTGCGCAACAGGGACAGAAAGAGAGAGAGAGGAGCGCTCGACACCTGGCCCT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 533 GGTCCAACCTGCGCAACAGGGACAGAAAGAGAGAGAGAGAGGAGCGCTGGTCCACCTGGCCCT 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 241 CGGGTGTAGAGAGGCCCAATTGGACCACTGGTCCCGCCGAGAGCGTGGCGGCAAGGA 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 593 CGGGTGTAGAGAGGCCCAATTGGACCACTGGTCCCGCCGAGAGCGTGGCGGCAAGGA 652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 301 TCTAAGGCTCCAGGGGCCCCAAAGGCTCCCGTGGTCCCTGGGAAGCCCGGCCCTCAG 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 653 TCCAAAGGCTCAGAGGTTCCTCAAGAGGATCTCTGGGTGTCCTCCAGGAGCGCTGGCCCTCAA 712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 361 GCGCCAGTGGGAGCCCAAGGCCCGCCCGCCACAGGCAAGAGAGGACTCCCGGCCCT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 713 GGACCTAGTGGGAGCCCAAGGACCAAGTCCACAGGAGGATGGACTCCCTGGCCCT 772
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 421 CAGGGCCCTCTGGTTCAGAGGACTTCAGGGCACCGTTGGGAGCCCTGGGGTGCTCGGA 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 773 CANGGCCCTCTCTGGCTT-CAGGGACTACAGGGCACTGTGGTGAGCTGTGAGTACTCTGGA 831
Oy 481 CCTGGGGACTG-CCAGGCTTGGCTGGGTACCAGGCATGCC---AGCCCCCAAGGCC 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 832 CCTCGGGGTTCGCCAGGCTTGCACAGGGTGCACAGCATGCCCTGGGCTTAAGGACCA 891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 537 CCGCGGCCCTCTCTGGCC 554
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 892 CTGGGCCCTCAAGGCC 909
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
BB248064 LOCUS
DEFINITION BB248064 RIKEN full-length enriched, 7 days neonate cerebellum Mus
musculus cDNA clone A730023E20 3', mRNA sequence.
VERSION BB248064.2 GI:16355610
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 638)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okada
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
JOURNAL On Jul 6, 2000 this sequence version replaced gi:8940810.
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Location/Qualifiers

```

```

source
1. .638
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730023E20"
/clone_lib="RIKEN full-length enriched, 7 days neonate
cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/notes="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAAGGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCGAGTTAATAATTAATCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."
BASE COUNT      212 a   134 c   146 g   146 t
ORIGIN

Query Match      36.0%; Score 369.4; DB 10; Length 638;
Best Local Similarity 84.5%; Pred. No. 3.4e-77;
Matches 415; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 536 CCOCGCCCTCTCGGCCATCAGGAGCGTGTGCCCTGGCCCTCGAGAAATGCT 655
Db 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1 CACCTGGCCCTCCAGGCCCTCAGGAGCAATGGAGGCATTGGCTCTCGAAGTGAACCAA 60

QY 596 CCCCAGGACCGGAGCAATGGCTGCCGCCTCACTGGAAGAACTTCACAGACAAATGCT 655
Db 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
61 CCCAGCATCAGAGTCAACGGATGTCGGCTCACTGGAAGAACTTCACAGATAATGCT 120

QY 656 ACTATTTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAAAGCTTTTCTGTGAAGACAAGT 715
Db 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
121 ACTATTTTTCATTTGAAAAAGAAATTTTGAAGATGCTAAGCTTTTCTGTGAAGACAAAT 180

QY 716 CTTACATCTTGTTCATTAACACTAGAGGAGACAGCAATGATGATAAAAACAGATGG 775
Db 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
181 CTTCCCATCTCGTTTCATTAACCTCAAGAGAAACAGCAATGATGATAAAAAGCATACCG 240

QY 776 TAGGAGAGAGAGCACTGGATCGGCCTCAAGACTCAGAGCGTGAAGTGAATGGAAGT 835
Db 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
241 TGGGAGAGAAAGCCATTTGATCGGCCTCAAGACTCAGACAGGAAAGCAATGGAAGT 300

QY 836 GGCTGGATGGACATCTCCAGACTTACAAAATTTGAAAGCTGGACAGCCGATTAACATGGG 895
Db 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
301 GGCTAGACGGTCACTCTGTTGATTACAAAACTGGAAGCTGGACAAACAGATAAATGGG 360

QY 896 GTCATGCCCATTGGCCAGGAGAGAACTGTGCTGGGTTGATTTATGCTGGCAGATGGAACG 955
Db 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
361 GCATGGCCCATGGCCAGGAGAGAACTGTGCTGGCTTGTATTTAGCGAGGACATGGAATG 420

QY 956 ATTTCGAATGTGAAGACGTCAAATAACTTTCATTCGAAAAAGACAGGAGACACTACTGT 1015
Db 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
421 ACTTCAGTGTGTAATCAATACTTCATTTGTGAGAGGAAAGGAGCGCATCATCAT 480

QY 1016 CATCTGCATTA 1026
Db 1 | | | | |
481 CATCCATATTA 491

```

UI-E-EJ0-ahq-c-05-0-UI.s2 UI-E-EJ0 Homo sapiens cDNA clone
 UI-E-EJ0-ahq-c-05-0-UI 3', mRNA sequence.
 BM676508
 BM676508.1 GI:18986404
 EST.

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 500)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
 Normalization and subtraction: two approaches to facilitate gene
 discovery
JOURNAL
 Genome Res. 6 (9), 791-806 (1996)
MEDLINE
 9704477
COMMENT
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hagaman
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

| | |
|---------------------|---|
| Location/Qualifiers | |
| 1..500 | /organism="Homo sapiens" |
| | /db_xref="taxon:9606" |
| | /clone="UI-E-EJ0-ahq-c-05-0-UI" |
| | /clone_lib="UI-E-EJ0" |
| | /tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid" |
| | /dev_stage="fetal and adult" |
| | /lab_host="DH10B (life Technologies) (Tl phage resistant)" |
| | /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-EJ0 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, ACATCAAGA ; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT; optic nerve, CCATTAAAGTC; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI). TAG_LIB=UI-E-EJ0 TAG_TISSUE=human eye TAG_SEQ=AATGCCGCAT" |
| BASE COUNT | 109 a 123 c 89 g 179 t |
| ORIGIN | |

Query Match 34.2%; Score 351.4; DB 14; Length 500;
 Best Local Similarity 99.7%; Pred. No. 5.9e-73;
 Matches 352; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 674 AAGNAATTTTGAGGACGAAGCTTTCTGTGAGACAAGCTCTTCACACTCTGGTTTCA 733
 | |||||
 Db 494 AGGAATTTTTGAGGTGCAAGCTTTTCTGTGAAGACAAGTCTTCACACTCTGGTTTCA 435
 | |||||
 QY 734 TAACACTAGAGGAACGAACATGGATAAAAAAACAGATGGTAGGAGAGAGAGCCACT 793


```

|||||
Db 434 TAAACACTAGACAGAACCAATGGATAAAAAACAAGATGTTAGGAGAGAGAGCCACT 375
|||||
Qy 794 GGATGGCTCACAGACTCAGAGCGTGAATAATGAATGAAGTGGTGGATGGAGCATCTC 853
|||||
Db 374 GGATGGCTCACAGACTCAGAGCGTGAATAATGAATGAAGTGGTGGATGGAGCATCTC 315
|||||
Qy 854 CAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTGTATGGCCATGGCCAG 913
|||||
Db 314 CAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTGTATGGCCATGGCCAG 255
|||||
Qy 914 GAGAAGCTGTCTGGTGTATTTATGCTGGCCAGTGGACGATTTCCATGTGAAGAGC 973
|||||
Db 254 GAGAAGCTGTCTGGTGTATTTATGCTGGCCAGTGGACGATTTCCATGTGAAGAGC 195
|||||
Qy 974 TCAATAAATTCATTTGCGAAAAGACAGGGAGACAGTACTGTCTATCTGCATTA 1026
|||||
Db 194 TCAATAAATTCATTTGCGAAAAGACAGGGAGACAGTACTGTCTATCTGCATTA 142
|||||

RESULT 12
LOCUS BM713891
DEFINITION UI-E-EJ0-ahq-c-05-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone
          UI-E-EJ0-ahq-c-05-0-UI 5', mRNA sequence.
ACCESSION BM713891
VERSION   1
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 507)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
           discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   9704477
COMMENT   Contact: Soares, MB
           Program for Rat Gene Discovery and Mapping
           University of Iowa
           451 Eckstein Medical Research Building Iowa City, IA 52242, USA
           Tel: 319 335 8250
           Fax: 319 335 9565
           Email: msources@blue.weeg.uiowa.edu
           Tissue Procurement: Dr. Gregg Hageman
           cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
           DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
           DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
           Clone Distribution: Researchers may obtain clones from Research
           Genetics (www.resgen.com).
           The following repetitive elements were found in this cDNA
           sequence: 443-498, >POLY_A#Simple_repeat
Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source           1..507
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-EJ0-ahq-c-05-0-UI"
                     /clone_lib="UI-E-EJ0"
                     /tissue_type="fetal eyes, lens, eye anterior segment,
                     optic nerve, retina, Retina Foveal and Macular, RPE and
                     Choroid"
                     /dev_stage="fetal and adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-EJ0 is a subtracted cDNA library constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT73-Pac

```

```

vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAAGA
; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT;
optic nerve, CCATTAAGTG; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
BASE COUNT 184 a 89 c 121 g 112 t 1 others
ORIGIN

Query Match      34.1%; Score 350; DB 14; Length 507;
Best Local Similarity 100.0%; Pred. No. 1.3e-72;
Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 677 AAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTTGTTTCATAA 736
|||||
Db 1 AAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTTGTTTCATAA 60
|||||
Qy 737 ACACCTAGAGAGGAACAGCAATGGATAAAAAACAGATGTTAGGAGAGAGAGCCACTGGA 796'
|||||
Db 61 ACACCTAGAGAGGAACAGCAATGGATAAAAAACAGATGTTAGGAGAGAGAGCCACTGGA 120'
|||||
Qy 797 TCGGCCTCACAGACTCAGAGCGTGAATAATGAATGAAGTGGTGGATGGAGCATCTCCAG 856
|||||
Db 121 TCGGCCTCACAGACTCAGAGCGTGAATAATGAATGAAGTGGTGGATGGAGCATCTCCAG 180
|||||
Qy 857 ACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTGTATGGCCATGGCCAGGAG 916
|||||
Db 181 ACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTGTATGGCCATGGCCAGGAG 240
|||||
Qy 917 AAGACTGTGCTGGTGTATTTATGCTGGGCAGTGGAAACGATTTCCTCAATGTGAAGAGCTCA 976
|||||
Db 241 AAGACTGTGCTGGTGTATTTATGCTGGGCAGTGGAAACGATTTCCTCAATGTGAAGAGCTCA 300
|||||
Qy 977 ATAACCTCATTTGCGAAAAGACAGGAGACAGTACTGTCTATCTGCATTA 1026
|||||
Db 301 ATAACCTCATTTGCGAAAAGACAGGAGACAGTACTGTCTATCTGCATTA 350
|||||

RESULT 13
LOCUS BM771366
DEFINITION UI-M-FI0-byu-g-09-0-UI.r1 NIH_BMAP_FI0 Mus musculus cDNA clone
          IMAGE:5702432 5', mRNA sequence.
ACCESSION BM771366
VERSION   1
KEYWORDS  EST.
SOURCE    house mouse.
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 808)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: csapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pyX-5.
           Location/Qualifiers
           1..808
           /organism="Mus musculus"
           /strain="C57BL/6"

```

```
/db_xref="taxon:10090"
/clone="IMAGE:5702432"
/clone_lib="NIH_BMAP_F10"
/tissue_type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="DH10B (T1 phage resistant)"
Note="Organ: Brain; Vector: ppx- Asc; Site_1: Ecor I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into ppx-Asc vector. The library tag is CAGCCACACAC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT 270 a 158 c 173 g 206 t 1 others
ORIGIN
Query Match 33.5%; Score 343.8; DB 14; Length 808;
Best Local Similarity 85.1%; Pred. No. 4.4e-71;
Matches 384; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 576 GGGCTGCGAGATGACCGCAACCCCGCAGGAGGACAAATGGCTGCCGCCCTCACTGGAA 635
D 1 GGGCTGCGAGATGACCGCAACCCCGCAGGAGGACAAATGGCTGCCGCCCTCACTGGAA 60
QY 636 GAATTCACAGCAAAATGCTACTATTTTCAGTTGAGGAAGAAATTTTGGAGATGCCAA 695
D 61 GAATTCACAGCAAAATGCTACTATTTTCATTTGGAAGAAAGAAATTTTGAAGATGCTAA 120
QY 696 GTTTTCTGTGAAGCAAGCTTTCACATCTGTTTTCATATAACACTAGAGAGAAACAGCA 755
D 121 GCTTTTCTGTGAAGCAAAATCTTCCATCTGCTTTTTCATATAACACTAGAGAGAAACAGCA 180
QY 756 ATGGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCTCACAGACTCAGA 815
D 181 ATGGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCTCACAGACTCAGA 240
QY 816 GCGTGAATAATGAATGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGGAAGC 875
D 241 ACAGGAACCGAATGGAAGTGGCTAGACGGGTCACTGTGTTATACAAAACTGGGAACG 300
QY 876 TGGACAGCCGGATTAAGTGGGTCATGGCCATGGCCAGAGAGACTGTGCTGGTGTAT 935
D 301 TGGACAGCCAGATAACTGGGCGAGTGGCCATGGCCAGAGAGACTGTGCTGGTGTAT 360
QY 936 TTATGCTGGCAGTGAACGANTTCCAATGTGAAGACGTCATTAATTAATTTGCGAATA 995
D 361 TTACGAGAGACAGTGAATGACTTCCAGTGTGATGAATCAATAATCAATTCATTGTGAGAA 420
QY 996 AGACAGGAGAGACGACTGCTGCTGTCATTG 1026
D 421 GGAAAGGAGGAGGAGTACCATCATCATTA 451
RESULT 14
AA304251
LOCUS
DEFINITION
339 bp mRNA linear EST 18-APR-1997
Estr17006 Aorta endothelial cells, TNF alpha-treated Homo sapiens
cDNA 5' end similar to similar to Fc receptor for IgE, low affinity
, mRNA sequence.
ACCESSION
AA304251
VERSION
AA304251.1 GI:1956604
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

REFERENCE
AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 339)
Adams M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult
, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White
, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai C., Clayton, R.A.,
Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald
, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A.,
Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,
Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,
Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L.,
Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon
, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M., and
Venter, J.C.

TITLE
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)

JOURNAL
MEDLINE
COMMENT

Other ESTs: THC84598
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavetigr.org

For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES
source

1. 339
/organism="Homo sapiens"
/db_xref="ATCC (inhost):115883"
/db_xref="taxon:9606"
/clone_lib="Aorta endothelial cells, TNF alpha-treated"
/cell_type="endothelial cell"
/dev_stage="adult"
/note="Organ: aorta; Vector: pBluescript SK-; Site_1:
EcorI; Site_2: XhoI"

BASE COUNT 107 a 69 c 86 g 71 t 6 others
ORIGIN

Query Match 30.3%; Score 311; DB 9; Length 339;
Best Local Similarity 96.2%; Pred. No. 2.1e-63;
Matches 326; Conservative 0; Mismatches 11; Indels 2; Gaps 1;

QY 573 CTGCGCCCTGCGAGATGACCGCAACCCCGCAGGAGGACAAATGGCTGCCGCCCTCACTG 632
D 1 CTGCGCCCTGCGAGATGACCGCAACCCCGCAGGAGGACAAATGGCTGCCGCCCTCACTG 60
QY 633 GAAGAACTTCACAGCAAAATGCTACTATTTTCAGTTGAGAAGAAATTTTGGAGATGC 692
D 61 GAAGAACTTCACAGCAAAATGCTACTATTTTCAGTTGAGAAGAAATTTTGGAGATGC 120
QY 693 AAAGCTTTTCTGTGAAGACAAGCTTTCACATCTGTTTTCATAAACAACACTAGAGAGAA 752
D 121 AAAGCTTTTCTGTGAAGACAAGCTTTCACATCTGTTTTCATAAACAACACTAGAGAGAA 180
QY 753 GCAATGGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCTCACAGACTC 812
D 181 GCAATGGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCTCACAGACTC 240
QY 813 AGAGCGTGAAATGAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGAA 872
D 241 AGAGCGTGAAATGAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGAA 300
QY 873 AGCTGGACAGCCGGAT--AACTGGGGTCTATGCCATGGG 909

```
Qy 918 AGACTGTGCTGGGTTGATTATGCTGG 944
    |||||
Db 316 AGACTGTGCTGGGTTGATTATGCTGG 342

Search completed: March 21, 2003, 08:08:21
Job time : 1357.02 secs
```


| Result No. | Query | | | Length | DB | ID | Description |
|------------|--------|-------|------|--------|----------|--------------------|-------------|
| | Score | Match | % | | | | |
| 1 | 1026 | 100.0 | 2024 | 21 | AAA07697 | Human collectin re | |
| 2 | 1026 | 100.0 | 2262 | 24 | ABA97932 | Human scavenger re | |
| 3 | 1026 | 100.0 | 2628 | 22 | AAH43036 | Nucleotide sequenc | |
| 4 | 1024.4 | 99.8 | 2005 | 24 | ABL95574 | Human angiogenesis | |
| 5 | 1024.4 | 99.8 | 2005 | 24 | ABL88085 | Human PRO7223 cDNA | |
| 6 | 1024.4 | 99.8 | 2181 | 22 | AAI60628 | Human polynucleoti | |
| 7 | 1024.4 | 99.8 | 2318 | 22 | AAI58842 | Human polynucleoti | |
| 8 | 1024.4 | 99.8 | 2641 | 22 | AAC66903 | Human EXMAD-14 cod | |
| 9 | 1024.4 | 99.8 | 2929 | 24 | ABQ92072 | Human polynucleoti | |

FT misc_feature /note= "fragment specifically claimed in claim 8"
FT 670..1695
FT /*tag= g
FT /note= "fragment specifically claimed in claim 5"
FT 685..738
FT /*tag= h
FT /note= "fragment specifically claimed in claim 7"
FT 730..738
FT /*tag= i
FT /note= "fragment specifically claimed in claim 7"
FT 739..1695
FT /*tag= j
FT /note= "fragment specifically claimed in claim 6"
FT 1696..2024
FT /*tag= k
FT /note= "fragment specifically claimed in claim 9"
XX
XX WO200011161-A1.
PN
XX
XX 02-MAR-2000.
PD
XX
XX 24-AUG-1999; 99WO-JP04552.
PF
XX
XX 24-AUG-1998; 98JP-0237611.
PR
XX
XX (FUSO) FUSO PHARM IND LTD.
PA
XX
XX Wakamiya N;
PI
XX
XX WPI: 2000-224696/19.
DR
XX P-PSDB; AAY7985.
DR
XX
XX New collectin of human origin having antibacterial and antiviral
PT activity, and gene encoding it useful for production of transgenic
PT animals and of antibodies for screening potential drug molecules -
PT
XX
XX Claim 5-9; Page 87-90; 106pp; Japanese.
PS
XX
XX The invention relates to polynucleotides encoding a new collectin of
CC human origin. The collectin can be used as an antibacterial and antiviral
CC agent and for screening potential drug molecules. The new collectin can
CC be produced by standard recombinant methodology. The present sequence
CC represents a DNA encoding the human collectin.
XX
XX Sequence 2024 BP; 595 A; 524 C; 506 G; 399 T; 0 other;
SQ
Query Match 100.0%; Score 1026; DB 21; Length 2024;
Best Local Similarity 100.0%; Pred. No. 2.4e-251;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCCAACAGATTGATGAGGTCGAGGTTAGACACTGAAAGTAGCCCAACTTATCAGTGATT 60
Db 670 ATGCCAACAGATTGATGAGGTCGAGGTTAGACACTGAAAGTAGCCCAACTTATCAGTGATT 729
QY 61 ATGGAGAAATGAAGTAGTAGTACCTCCAAAGCATGTCAGCTCATCAAGAAATTTTACAATA 120
Db 730 ATGGAGAAATGAAGTAGTAGTACCTCCAAAGCATGTCAGCTCATCAAGAAATTTTACAATA 789
QY 121 CTACAGGTCCACCGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCACGAGGCCCT 180
Db 790 CTACAGGTCCACCGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCACGAGGCCCT 849
QY 181 GCGCCCACTGCCAACAGGAGCAGAAAGGAGAGAGGGGGAGCCCTGGACCACTGGCCCT 240
Db 850 GCGCCCACTGCCAACAGGAGCAGAAAGGAGAGAGGGGGAGCCCTGGACCACTGGCCCT 909
QY 241 GCGGGTGAGAGAGGCCCAATTGGACAGCTGTGTCCTCCCGGAGAGCGTGGCGGCAAGGA 300
Db 910 GCGGGTGAGAGAGGCCCAATTGGACAGCTGTGTCCTCCCGGAGAGCGTGGCGGCAAGGA 969
QY 301 TCTAAGGCTCCACGGGCCCCAAAGGCTCCCGTGGTTCCTCGGGAAGCCCGGCTCAG 360
Db 970 TCTAAGGCTCCACGGGCCCCAAAGGCTCCCGTGGTTCCTCGGGAAGCCCGGCTCAG 1029

QY 361 GCGCCCACTGGGAGCCAGGCCCGCCCGGCCCCACAGGCCCAAGAGGGAGCTCCCGGCCCT 420
Db 1030 GCGCCCACTGGGAGCCAGGCCCGCCCGGCCCCACAGGCCCAAGAGGGAGCTCCCGGCCCT 1089
QY 421 CAGGGCCCTCCCTGGCTTCCAGGGACTTCAGGGCACCCTGTGGGGAGCCCTGGCTGGA 480
Db 1090 CAGGGCCCTCCCTGGCTTCCAGGGACTTCAGGGCACCCTGTGGGGAGCCCTGGCTGGA 1149
QY 481 CCTCGGGGACTGCCAGGCTTGGCTGGGGTACCAAGGATGCCAGGCCCGCCCGGCCCC 540
Db 1150 CCTCGGGGACTGCCAGGCTTGGCTGGGGTACCAAGGATGCCAGGCCCGCCCGGCCCC 1209
QY 541 GCGCCCTCCCTGGCCCATCAGAGCGGTGTGCCCCCTGGCCCTGCAGAATGAGCCCAAGCC 600
Db 1210 GCGCCCTCCCTGGCCCATCAGAGCGGTGTGCCCCCTGGCCCTGCAGAATGAGCCCAAGCC 1269
QY 601 GCACCGGAGGACAATGGCTGGCCCGCCCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 660
Db 1270 GCACCGGAGGACAATGGCTGGCCCGCCCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 1329
QY 661 TTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCTTTTCTGTCAAGACAAGTCTTCA 720
Db 1330 TTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCTTTTCTGTCAAGACAAGTCTTCA 1389
QY 721 CATCTTGTGTTTATCAAACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGG 780
Db 1390 CATCTTGTGTTTATCAAACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGG 1449
QY 781 AGAGAGAGCCACTGGATGGCCCTCAGACTCAGAGCTGAAAATGAATGGAAGTGGCTG 840
Db 1450 AGAGAGAGCCACTGGATGGCCCTCAGACTCAGAGCTGAAAATGAATGGAAGTGGCTG 1509
QY 841 GATGGGACATCTCCAGACTACAAAATTCGAAAGCTGGACAGCCGGATAACTGGGGTCAT 900
Db 1510 GATGGGACATCTCCAGACTACAAAATTCGAAAGCTGGACAGCCGGATAACTGGGGTCAT 1569
QY 901 GCGCATGGCCAGGAGAGACTGTGCTGGGTTGATTTATGCTGGGAGTGGACGATTTTC 960
Db 1570 GCGCATGGCCAGGAGAGACTGTGCTGGGTTGATTTATGCTGGGAGTGGACGATTTTC 1629
QY 961 CAATGTGAAGAGCTCAATAACTTCATTTCCGAAAAGACAGGAGACACTGTCATCT 1020
Db 1630 CAATGTGAAGAGCTCAATAACTTCATTTCCGAAAAGACAGGAGACACTGTCATCT 1689
QY 1021 GCATTA 1026
Db 1690 GCATTA 1695
RESULT 2
ABA97932
ID ABA97932 standard; cDNA; 2262 BP.
XX ABA97932;
AC ABA97932;
XX
XX 25-APR-2002 (first entry)
XX
DE Human scavenger receptor-like protein encoding cDNA SEQ ID NO 1.
XX
XX Human; scavenger; receptor; diagnosis; treatment; autoimmune disease;
XX rheumatoid arthritis; gene; ss.
XX Homo sapiens.
OS
XX
XX Location/Qualifiers
FH Key 33..2262
FT CDS /*tag= a
FT /product= "scavenger receptor-like protein"
XX JP2001340089-A.
XX
XX 11-DEC-2001.
PD

XX 08-DEC-2000; 2000JP-0375066.
PF XX
PR XX
XX 27-MAR-2000; 2000JP-0090772.
PR XX
XX (SHIO) SHIONOGI & CO LTD.
PA XX
XX WPI; 2002-144965/19.
DR XX
DR P-PSDB; ABB08642.
XX XX
PT New scavenger receptor-like protein for diagnosis, prevention and
PT treatment of autoimmune disease, such as rheumatoid arthritis -
XX Claim 3; Fig 1; 38pp; Japanese.
PS XX
XX The invention relates to a human scavenger receptor-like protein. The
CC protein is useful as a target molecule for diagnosis, prevention and
CC treatment of autoimmune diseases such as rheumatoid arthritis.
XX XX
SQ Sequence 2262 BP; 670 A; 576 C; 581 G; 435 T; 0 other;

Query Match 100.0%; Score 1026; DB 24; Length 2262;
Best Local Similarity 100.0%; Pred. No. 2.5e-251;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCAACAAGATTTGATGAGGCTGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
Db 1234 ATGCAACAAGATTTGATGAGGCTGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 1293

Qy 61 ATGGAGAATAGAGCTAGTACTCAAGCATGCTAGCTCATCAAGAAATTTACAATA 120
Db 1294 ATGGAGAATAGAGCTAGTACTCAAGCATGCTAGCTCATCAAGAAATTTACAATA 1353

Qy 121 CTAAAGTCTCCAGGGCCCGAGGGTCCAAAGAGTGACAGATCCACAGGACCCCT 180
Db 1354 CTAAAGTCTCCAGGGCCCGAGGGTCCAAAGAGTGACAGATCCACAGGACCCCT 1413

Qy 181 GCGCCAACTGTCACAAAGGACAGAAAGGAGAGAGGGGAGCCCTGGACCACTTGGCCCT 240
Db 1414 GCGCCAACTGTCACAAAGGACAGAAAGGAGAGAGGGGAGCCCTGGACCACTTGGCCCT 1473

Qy 241 CGGGTGAGAGGCCCAATTTGGACAGCTGCTCCCCCGAGAGCGTGGCGGCAAGGA 300
Db 1474 CGGGTGAGAGGCCCAATTTGGACAGCTGCTCCCCCGAGAGCGTGGCGGCAAGGA 1533

Qy 301 TCTAAGGCTCCAGGGCCCGCAAGGCTCCGCTGCTCCCTGGGAAGCCGCGCCCTCAG 360
Db 1534 TCTAAGGCTCCAGGGCCCGCAAGGCTCCGCTGCTCCCTGGGAAGCCGCGCCCTCAG 1593

Qy 361 GCGCCCACTGGGGACCCAGGCCCCCGGGCCCAAGAGAGGACTCCCGGGCCCT 420
Db 1594 GCGCCCACTGGGGACCCAGGCCCCCGGGCCCAAGAGAGGACTCCCGGGCCCT 1653

Qy 421 CAGGGCCCTCTGGCTTCCAGGAGCTTCAGGACACCGTTGGGGAGCGTGGGGTGGCTGGA 480
Db 1654 CAGGGCCCTCTGGCTTCCAGGAGCTTCAGGACACCGTTGGGGAGCGTGGGGTGGCTGGA 1713

Qy 481 CCTGGGGAGTCCAGGCTTGCCTGGGGTACAGGATGCGAGGCCCAAGGCCGCCGCC 540
Db 1714 CCTGGGGAGTCCAGGCTTGCCTGGGGTACAGGATGCGAGGCCCAAGGCCGCCGCC 1773

Qy 541 GCGCCCTCTGGCCCATCAGGAGCGTGGTGCCTTGGCCCTGCAGATGAGCCCAACCCCG 600
Db 1774 GCGCCCTCTGGCCCATCAGGAGCGTGGTGCCTTGGCCCTGCAGATGAGCCCAACCCCG 1833

Qy 601 GCACGGAGGACATGGTCCCGCTCAGTGGAGAACTTCACAGACAAATGCTACTAT 660
Db 1834 GCACGGAGGACATGGTCCCGCTCAGTGGAGAACTTCACAGACAAATGCTACTAT 1893

Qy 661 TTTTCAGTTGAGAAGAAATTTTGGAGATCAAGCTTTTCTGTGAGACAAAGTCTTCA 720
Db 1894 TTTTCAGTTGAGAAGAAATTTTGGAGATCAAGCTTTTCTGTGAGACAAAGTCTTCA 1953

Qy 721 CATCTTGTTCATAAAACACTAGAGAGAACAGCAATGGATAAAAAACAGATGTAGGG 780
Db 1954 CATCTTGTTCATAAAACACTAGAGAGAACAGCAATGGATAAAAAACAGATGTAGGG 2013

Qy 781 AGAGAGAGCCCACTGGATCGGCTCACAGACTCAGAGCGTGAATGGAATGGAGTGGCTG 840
Db 2014 AGAGAGAGCCCACTGGATCGGCTCACAGACTCAGAGCGTGAATGGAATGGAGTGGCTG 2073

Qy 841 GATGGACATCTCCAGACTACAAAAATTTGAAAGCTGCACAGCCGAGATAACTTGGGGTCTAT 900
Db 2074 GATGGACATCTCCAGACTACAAAAATTTGAAAGCTGCACAGCCGAGATAACTTGGGGTCTAT 2133

Qy 901 GGCCATGGCCAGGAGAGACTGTCTGGGTTGATTTATGCTGGCAGTGGAAACGATTTC 960
Db 2134 GGCCATGGCCAGGAGAGACTGTCTGGGTTGATTTATGCTGGCAGTGGAAACGATTTC 2193

Qy 961 CAATGTGAAGACGTCATTAACCTTCATTTGCGAAAAAGACAGGAGACAGTACTGTCTCT 1020
Db 2194 CAATGTGAAGACGTCATTAACCTTCATTTGCGAAAAAGACAGGAGACAGTACTGTCTCT 2253

Qy 1021 GCATTA 1026
Db 2254 GCATTA 2259

RESULT 3
AAH43036
ID AAH43036 standard; DNA; 2628 BP.
XX AC AAH43036;
XX DT 15-OCT-2001 (first entry)
XX Nucleotide sequence of a human scavenger receptor.
XX DE
XX Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
KW low density lipoprotein; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 74..2302
FT FT /*tag= a
FT FT /product= "scavenger receptor"

XX WO200159107-A1.
XX PD 16-AUG-2001.
XX PF 08-FEB-2001; 2001WO-JP00874.
XX PR 14-FEB-2000; 2000JP-0035155.
XX PR 10-OCT-2000; 2000JP-0309068.
XX PA (FUSO) FUSO PHARM IND LTD.
XX PI Wakamiya N;
XX WPI; 2001-497076/54.
DR P-PSDB; AAG63346.
XX New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
PT useful for treatment and diagnosis of diseases associated with oxidized
PT low-density lipoprotein accumulation
XX Claim 2; Page 79-84; 118pp; Japanese.
XX The present sequence encodes a human scavenger receptor, designated
CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
CC are useful in clarifying the functions of macrophages and basal
CC immunity. They are also useful in the treatment, prevention, diagnosis
CC and investigation of diseases such as arteriosclerosis, diabetic

PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220664P.
PR 28-JUL-2000; 2000WO-US20710.
PR 02-AUG-2000; 2000US-222695P.
PR 17-AUG-2000; 2000US-064365P.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 07-SEP-2000; 2000US-230978P.
PR 18-SEP-2000; 2000US-0664610.
PR 18-SEP-2000; 2000US-0665350.
PR 24-OCT-2000; 2000US-242922P.
PR 08-NOV-2000; 2000US-0709238.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 22-JAN-2001; 2001US-0767609.
PR 28-FEB-2001; 2001US-0796498.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 25-MAY-2001; 2001WO-US17092.
PR 30-MAY-2001; 2001US-0870574.
PR 30-MAY-2001; 2001WO-US17443.
PR 01-JUN-2001; 2001WO-US17800.
XX
XX (GETH) GENENTECH INC.
XX
XX Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
XX Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
XX Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
XX
XX WPI: 2002-090516/12.
XX P-PSDB; ABB84830.
XX
XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,
XX useful in diagnosis and treatment of cardiovascular (e.g. myocardial
XX infarction), endothelial or angiogenic disorders in a mammal -
XX
XX Claim 2; Fig 27; 565pp; English.
XX
XX ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to
XX ABB85003. The PRO proteins and polynucleotides have cardiant, cytostatic,
XX antiangiogenic, hypotensive, vulnerary and antiarteriosclerotic
XX activities, and can be used in gene therapy. The PRO polynucleotides,
XX proteins, agonists and antagonists are useful for treating or diagnosing
XX a cardiovascular, endothelial or angiogenic disorder in a mammal,
XX e.g. cardiac hypertrophy, trauma, cancer, age-related macular
XX degeneration, atherosclerosis, hypertension, arterial restenosis,
XX rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis,
XX lymphangitis, tumour angiogenesis (such as breast carcinoma and liver
XX carcinoma) and wound healing. The PRO polynucleotides have applications
XX in molecular biology, including use as hybridisation probes, and in
XX chromosome and gene mapping. ABL88259 to ABL88267 represent primers and
XX probes used in the exemplification of the present invention.
XX
XX Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;
XX
XX Query Match 99.8%; Score.1024.4; DB 24; Length 2005;
XX Best Local Similarity 99.9%; Pred. No. 6.1e-251;
XX Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 ATGCAACAGATTTCATGAGGTCGAGGTTAGACTGAAGTAGCCCAACTATCAGTGATT 60
XX
XX 478 ATGCAACAGATTTCATGAGGTCGAGGTTAGACTGAAGTAGCCCAACTATCAGTGATT 537
XX
XX

QY 61 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGTCAGCTCATCAAGAATTTTACAATA 120
DB 538 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGTCAGCTCATCAAGAATTTTACAATA 597
QY 121 CTACAAGGTCACCGGGCCCCAGGGGTCCAAAGAGTGACAGAGGATCCAGGAGCCCCCT 180
DB 598 CTACAAGGTCACCGGGCCCCAGGGGTCCAAAGAGTGACAGAGGATCCAGGAGCCCCCT 657
QY 181 GGCCCAACTGGCAACAAGGAG 240
DB 658 GGCCCAACTGGCAACAAGGAG 717
QY 241 GCGGGTGAG 300
DB 718 GCGGGTGAG 777
QY 301 TCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTCCCTGGGAGAGAGAGAGAGAGAGAGAG 360
DB 778 TCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTCCCTGGGAGAGAGAGAGAGAGAGAG 837
QY 361 GGCCCACTGGGGAG 420
DB 838 GGCCCACTGGGGAG 897
QY 421 CAGGGCCCTCTCTGGGCTTCAGGGGAGCTTCAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 898 CAGGGCCCTCTCTGGGCTTCAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 957
QY 481 CCGGGGAG 540
DB 958 CCGGGGAG 1017
QY 541 GGCCCTCTCTGGGCTTCAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 1018 GGCCCTCTCTGGGCTTCAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1077
QY 601 GCACGGGAG 660
DB 1078 GCACGGGAG 1137
QY 661 TTTTCAGTTGAG 720
DB 1138 TTTTCAGTTGAG 1197
QY 721 CATCTTGTGTTTTCATAAACACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 1198 CATCTTGTGTTTTCATAAACACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
QY 781 AG 840
DB 1258 AG 1317
QY 841 GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGAGTAACCTGGGGTCAAT 900
DB 1318 GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGAGTAACCTGGGGTCAAT 1377
QY 901 GGCCATGGGCCAGGAG 960
DB 1378 GGCCATGGGCCAGGAG 1437
QY 961 CAATCTGAAG 1020
DB 1438 CAATCTGAAG 1497
QY 1021 GCATTA 1026
DB 1498 GCATTA 1503
RESULT 6
AAI60628
ID AAI60628 standard; cDNA; 2181 BP.

| | | |
|----|---|--|
| XX | 22-OCT-2001 (first entry) | |
| DT | Human polynucleotide SEQ ID NO 1045. | |
| DE | Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; | |
| XX | peripheral nervous system; neuropathy; central nervous system; CNS; | |
| XX | Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; | |
| KW | amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; | |
| KW | chemokinetic; thrombolytic; drug screening; arthritis; inflammation; | |
| KW | leukaemia; ss. | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| PN | WO200153312-A1. | |
| XX | | |
| PD | 26-JUL-2001. | |
| XX | | |
| PF | 26-DEC-2000; 2000WO-US34263. | |
| XX | | |
| PR | 21-JAN-2000; 2000US-0488725. | |
| PR | 25-APR-2000; 2000US-0552317. | |
| PR | 09-JUL-2000; 2000US-0598042. | |
| PR | 19-AUG-2000; 2000US-0620312. | |
| PR | 03-AUG-2000; 2000US-0653450. | |
| PR | 14-SEP-2000; 2000US-0662191. | |
| PR | 19-OCT-2000; 2000US-0693036. | |
| PR | 29-NOV-2000; 2000US-0727344. | |
| XX | | |
| PA | (HYSE-) HYSEQ INC. | |
| XX | | |
| PI | Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; | |
| PI | Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J; | |
| PI | Zhao QA, Zhou P, Goodrich R, Drmanac RT; | |
| XX | | |
| DR | WPI: 2001-442253/47. | |
| DR | P-PSDB; AAM39686. | |
| XX | | |
| PT | Novel nucleic acids and polypeptides, useful for treating disorders | |
| PT | such as central nervous system injuries - | |
| XX | | |
| PS | Claim 1; SEQ ID NO 1045; 10078pp; English. | |
| XX | | |
| CC | The invention relates to human nucleic acids (AA157798-AA161369) and | |
| CC | the encoded polypeptides (AAM38642-AAM42213) with nootropic, | |
| CC | immunosuppressant and cytostatic activity. The polynucleotides are useful | |
| CC | in gene therapy. A composition containing a polypeptide or polynucleotide | |
| CC | of the invention may be used to treat diseases of the peripheral nervous | |
| CC | system, such as peripheral nervous injuries, peripheral neuropathy and | |
| CC | localised neuropathies and central nervous system diseases, such as | |
| CC | Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic | |
| CC | lateral sclerosis, and Shy-Drager Syndrome. Other uses include the | |
| CC | utilisation of the activities such as: immune system suppression, | |
| CC | Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic | |
| CC | and thrombolytic activity, cancer diagnosis and therapy, drug screening, | |
| CC | assays for receptor activity, arthritis and inflammation, leukaemias and | |
| CC | C.N.S disorders. | |
| CC | Note: The sequence data for this patent did not form part of the printed | |
| CC | specification. | |
| XX | | |
| SQ | Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other; | |
| | Query Match 99.8%; Score 1024.4; DB 22; Length 2318; | |
| | Best Local Similarity 99.9%; Pred. No. 6.4e-251; | |
| | Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | |
| QY | 1 ATGCACACAGATTGATGAGTCGAGGTTAGACACTGAAGTAGCCCAACTTATCAGTGATT 60 | |
| DB | | |
| DB | 663 ATGCACACAGATTGATGAGTCGAGGTTAGACACTGAAGTAGCCCAACTTATCAGTGATT 722 | |
| QY | 61 ATGGGAAGAATGAAGCTAGTAGCTCCAGCATGTCAGCTCATCAAGAATTTTCAATA 120 | |
| DB | | |
| DB | 723 ATGGGAAGAATGAAGCTAGTAGCTCCAGCATGTCAGCTCATCAAGAATTTTCAATA 782 | |
| | | |

| | | |
|----|---|--|
| QY | 121 CTACAAGGTCCACCGGGCCCAAGGGTCCAAGAGGTGACAGAGGATCCACGAGGACCCCT 180 | |
| DB | | |
| DB | 783 CTACAAGGTCCACCGGGCCCAAGGGTCCAAGAGGTGACAGAGGATCCACGAGGACCCCT 842 | |
| QY | 181 GCCCAACTGGCAACAAGGACAGAAAAGAGAGAGAGGGGAGCCCTGGACCACTCGGCCCT 240 | |
| DB | | |
| DB | 843 GCCCAACTGGCAACAAGGACAGAAAAGAGAGAGAGGGGAGCCCTGGACCACTCGGCCCT 902 | |
| QY | 241 GCGGGTGAGAGAGGCCCAATTTGGACCACTGCTGCCCGGAGAGCGTGGCGGCAAGA 300 | |
| DB | | |
| DB | 903 GCGGGTGAGAGAGGCCCAATTTGGACCACTGCTGCCCGGAGAGCGTGGCGGCAAGA 962 | |
| QY | 301 TCTAAAGGCTCCACGAGGCCCAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGCCCTCAG 360 | |
| DB | | |
| DB | 963 TCTAAAGGCTCCACGAGGCCCAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGCCCTCAG 1022 | |
| QY | 361 GGGCCCACTGGGGACCCAGGCCCCCGGGCCCAACAGGCAAGAGGGACTTCCCGGCCCT 420 | |
| DB | | |
| DB | 1023 GGGCCCACTGGGGACCCAGGCCCCCGGGCCCAACAGGCAAGAGGGACTTCCCGGCCCT 1082 | |
| QY | 421 CAGGGCCCTCTGGGCTTCAGGGACTTCAGGSCACCGTTGGGGAGCCCTGGGGTGCCTGA 480 | |
| DB | | |
| DB | 1083 CAGGGCCCTCTGGGCTTCAGGGACTTCAGGGACCGTTGGGGAGCCCTGGGGTGCCTGA 1142 | |
| QY | 481 CCTCGGGACTGCCAGGCTTGGCTGGGGTACCAGGCTATGCCAGGCCCAAGGGCCCCCCC 540 | |
| DB | | |
| DB | 1143 CCTCGGGACTGCCAGGCTTGGCTGGGGTACCAGGCTATGCCAGGCCCAAGGGCCCCCCC 1202 | |
| QY | 541 GGGCCCTCTGGGCTATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCACCCCG 600 | |
| DB | | |
| DB | 1203 GGGCCCTCTGGGCTATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCACCCCG 1262 | |
| QY | 601 GCACCGGAGGACAATGGCTGCCCGCTCAGTGAAGAACTTCACAGACAAATGCTACTAT 660 | |
| DB | | |
| DB | 1263 GCACCGGAGGACAATGGCTGCCCGCTCAGTGAAGAACTTCACAGACAAATGCTACTAT 1322 | |
| QY | 661 TTTTCAGTTGAGAAAATAATTTTGGAGTGCAGGCTTTCTGTGAAGACAAAGTCTTCA 720 | |
| DB | | |
| DB | 1323 TTTTCAGTTGAGAAAATAATTTTGGAGTGCAGGCTTTCTGTGAAGACAAAGTCTTCA 1382 | |
| QY | 721 CATCTGTTTTCATAAAGCTAGAGGAGCAAGCAATGGATAAAGAAACAGATGGTAGG 780 | |
| DB | | |
| DB | 1383 CATCTGTTTTCATAAAGCTAGAGGAGCAAGCAATGGATAAAGAAACAGATGGTAGG 1442 | |
| QY | 781 AGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAAGTGAAGTGGCTG 840 | |
| DB | | |
| DB | 1443 AGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAAGTGAAGTGGCTG 1502 | |
| QY | 841 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTCA 900 | |
| DB | | |
| DB | 1503 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTCA 1562 | |
| QY | 901 GGGCATGGCCAGGAGAGACTGCTGGGTTGATTATGCTGGGAGTGAAGCAATTC 960 | |
| DB | | |
| DB | 1563 GGGCATGGCCAGGAGAGACTGCTGGGTTGATTATGCTGGGAGTGAAGCAATTC 1622 | |
| QY | 961 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAGAGAGGAGACAGTACTGTCATCT 1020 | |
| DB | | |
| DB | 1623 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAGAGAGGAGACAGTACTGTCATCT 1682 | |
| QY | 1021 GCATTA 1026 | |
| DB | | |
| DB | 1683 GCATTA 1688 | |
| | | |
| | RESULT 8 | |
| | AAC66903 | |
| ID | AAC66903 standard; cDNA; 2641 BP. | |
| XX | | |
| AC | AAC66903; | |
| XX | | |
| DT | 27-MAR-2001 (first entry) | |

XX Human EXMAD-14 coding sequence SEQ ID NO: 39.
DE
XX
XX Extracellular matrix and adhesion-associated protein; EXMAD: cancer;
KW inflammation; reproductive disorder; cardiovascular disorder;
KW immune disorder; musculoskeletal disorder; developmental disorder;
KW gastrointestinal disorder; cell proliferation disorder; ss.
XX
OS Homo sapiens.
XX
XX WO20068380-A2.
PN
XX 16-NOV-2000.
PD
XX 10-MAY-2000; 2000WO-US12811.
XX
XX 11-MAY-1999; 99US-0133643.
PR
XX 23-AUG-1999; 99US-0150409.
PR
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DAM;
PI Azimzai Y;
PI
XX
XX WPI: 2001-007395/01.
DR
XX P-PSDB: AAB27236.
DR
XX
XX Isolated polynucleotide encoding extracellular matrix or
PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
PT preventing disorders associated with expression of EXMAD such as
PT proliferative, immune and genetic disorders -
XX
XX Claim 4; Page 121-122; 129pp; English.
PS
XX
XX The present invention provides the protein and coding sequences for 25
CC novel extracellular matrix and adhesion-associated proteins (EXMADs).
CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
CC useful in the prevention and treatment of cancers, cell proliferation,
CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
CC gastrointestinal disorders and inflammation.
XX
XX Sequence 2641 BP; 823 A; 628 C; 608 G; 582 T; 0 other;
Query Match 99.8%; Score 1024.4; DB 22; Length 2641;
Best Local Similarity 99.9%; Pred. No. 6.8e-251;
Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAACAGATTTGATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
DB
DB 912 ATGCAACAGATTTGATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 971
QY 61 ATGGAAGAAATGAAGCTAGTACTTCCAAAGCATGCTCAGCTCATCAAGAATTTTACAATA 120
DB
DB 972 ATGGAAGAAATGAAGCTAGTACTTCCAAAGCATGCTCAGCTCATCAAGAATTTTACAATA 1031
QY 121 CTACAAGTCCACCGGGCCCGCCAGAGGTCACAGAGTCCAGAGGATCCAGGAGCCCT 180
DB
DB 1032 CTACAAGTCCACCGGGCCCGCCAGAGGTCACAGAGTCCAGAGGATCCAGGAGCCCT 1091
QY 181 GCGCCAACTGGCAACAGGGACAGAAAGAGAGAGAGGGGAGCCCTGGACCTGGCCCT 240
DB
DB 1092 GCGCCAACTGGCAACAGGGACAGAAAGAGAGAGAGGGGAGCCCTGGACCTGGCCCT 1151
QY 241 GCGGCTGAGAGAGGCCCCAATTTGACCACTGCTCCCGGGAGAGCGTGGCGGCAAGGA 300
DB
DB 1152 GCGGCTGAGAGAGGCCCCAATTTGACCACTGCTCCCGGGAGAGCGTGGCGGCAAGGA 1211
QY 301 TCTAAGGCTCCAGGGCCCCCAAGGCTCCCGTGGTCCCTGGGAAGCCCGGCCCTCAG 360
DB
DB 1212 TCTAAGGCTCCAGGGCCCCCAAGGCTCCCGTGGTCCCTGGGAAGCCCGGCCCTCAG 1271

QY 361 GGGCCAGTGGGAGCCAGGCCCCCGGGCCACCAGGCAAAAGAGGACTCCCCGGCCCT 420
DB
DB 1272 GGGCCAGTGGGAGCCAGGCCCCCGGGCCACCAGGCAAAAGAGGACTCCCCGGCCCT 1331
QY 421 CAGGCCCCCTCTGGCTTCAGGAGCTTCAGGAGCCCTGGGGAGCGCTGGGGTGGCTTGA 480
DB
DB 1332 CAGGCCCCCTCTGGCTTCAGGAGCTTCAGGAGCCCTGGGGAGCGCTGGGGTGGCTTGA 1391
QY 481 CCTCGGGAGCTGCCAGGCTTGCTGGGTACCCAGGATGCCAGGCCCAAGAGGCCCCGCC 540
DB
DB 1392 CCTCGGGAGCTGCCAGGCTTGCTGGGTACCCAGGATGCCAGGCCCAAGAGGCCCCGCC 1451
QY 541 GGGCCCTCTGGCCCCATCAGGAGCGGTGGTGGCCCCCTGCAGAAATGAGCAACCCCG 600
DB
DB 1452 GGGCCCTCTGGCCCCATCAGGAGCGGTGGTGGCCCCCTGCAGAAATGAGCAACCCCG 1511
QY 601 GCACCGGAGGACATGGCTGCCGCCCTCAGTGGAGAACTTCACAGACAAATGCTACTAT 560
DB
DB 1512 GCACCGGAGGACATGGCTGCCGCCCTCAGTGGAGAACTTCACAGACAAATGCTACTAT 1571
QY 661 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAAAGCTTTTCTGTGAACAGACTTCTCA 720
DB
DB 1572 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAAAGCTTTTCTGTGAACAGACTTCTCA 1631
QY 721 CATCTTGTTCATAAACACACTAGAGAGGAACAGCAATGATGATGATGATGATGATGATG 780
DB
DB 1632 CATCTTGTTCATAAACACACTAGAGAGGAACAGCAATGATGATGATGATGATGATGATG 1691
QY 781 AGAGAGCCACTGATGCGCTCAGAGACTCAGAGCGTGAAGTGAAGTGAAGTGAAGTGAAG 840
DB
DB 1692 AGAGAGCCACTGATGCGCTCAGAGACTCAGAGCGTGAAGTGAAGTGAAGTGAAGTGAAG 1751
QY 841 GATGGAGATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGAGTAACCTGGGCTAT 900
DB
DB 1752 GATGGAGATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGAGTAACCTGGGCTAT 1811
QY 901 GGGCATGGCCAGGAGAGAGACTGTCTGGTGTGATTTATGCTGGCAGTGGCAACGATTTTC 960
DB
DB 1812 GGGCATGGCCAGGAGAGAGACTGTCTGGTGTGATTTATGCTGGCAGTGGCAACGATTTTC 1871
QY 961 CAATGTGAAGAGCTCAATACTTCAATTTTCGAAAAGACAGGAGACAGTACTGATCT 1020
DB
DB 1872 CAATGTGAAGAGCTCAATACTTCAATTTTCGAAAAGACAGGAGACAGTACTGATCT 1931
QY 1021 GCATTA 1026
DB
DB 1932 GCATTA 1937
RESULT 9
ABQ92072
ID ABQ92072 standard; cDNA; 2929 BP.
XX
AC ABQ92072;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 69.
XX
KW Human; cytostatic; antirheumatic; antiarthritic; vulnary; analgesic;
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW antiulcer; fungicide; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW osteoporosis; severe combined immunodeficiency; SCID; infection;
KW multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
XX
XX Homo sapiens.
XX

| | | | |
|----|---|---------------------|--|
| XX | 23-MAR-1999 | (first entry) | |
| XX | Human secreted protein clone bv227_1 | coding sequence. | |
| DE | | | |
| XX | Secreted protein; human; nutritional supplements; immune stimulant; | | |
| XX | immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis; | | |
| KW | activin/inhibin; chemokinesis; haemostasis; thrombolysis; | | |
| KW | receptor/ligand activity; anti-inflammatory; tumour inhibitor; | | |
| KW | cadherin/tumour invasion suppressor; ds. | | |
| XX | | | |
| OS | Homo sapiens. | | |
| XX | | | |
| FH | Key | Location/Qualifiers | |
| FT | CDS | 67..693 | |
| FT | | /*tag= a | |
| XX | | | |
| PN | W09855614-A2. | | |
| XX | | | |
| PD | 10-DEC-1998. | | |
| XX | | | |
| PF | 01-JUN-1998; | 98WO-US11210. | |
| XX | | | |
| PR | 29-MAY-1998; | 98US-0087255. | |
| PR | 04-JUN-1997; | 97US-0868696. | |
| PR | 04-JUN-1997; | 97US-0868697. | |
| PR | 04-JUN-1997; | 97US-0868698. | |
| PR | 04-JUN-1997; | 97US-0868898. | |
| PR | 04-JUN-1997; | 97US-0868899. | |
| PR | 04-JUN-1997; | 97US-0868900. | |
| PR | 04-JUN-1997; | 97US-0869191. | |
| PR | 04-JUN-1997; | 97US-0869192. | |
| PR | 04-JUN-1997; | 97US-0869193. | |
| PR | 04-JUN-1997; | 97US-0869194. | |
| XX | | | |
| PA | (GMY) GENETICS INST INC. | | |
| XX | | | |
| PI | Agostino MJ, Fichtel K, Howes SH, Jacobs K, Lavallie ER; | | |
| PI | McCoy JM, Racie LA, Spaulding V, Treacy M; | | |
| XX | | | |
| DR | WPI: 1999-059912/05. | | |
| DR | P-FSDB: AAW73628. | | |
| XX | | | |
| PT | New polynucleotides encoding secreted human proteins - derived from | | |
| PT | human foetal brain, adult brain, foetal kidney, adult ovary, adult | | |
| PT | retina, adult placenta or adult uterus cDNA libraries | | |
| XX | | | |
| PS | Claim 26; Page 87-88; 127pp; English. | | |
| XX | | | |
| CC | This sequence encodes a human secreted protein of the invention. | | |
| CC | This DNA sequence was isolated from a human adult brain cDNA | | |
| CC | library, and was designated clone bv227.1. The DNAs and proteins | | |
| CC | are predicted to have biological activities which would make them | | |
| CC | suitable for treating, preventing or ameliorating medical conditions in | | |
| CC | humans and animals, although no supporting data is given. Suggested | | |
| CC | activities include nutritional sources or supplements, immune | | |
| CC | stimulating or suppressing activity, haematopoiesis regulating activity, | | |
| CC | tissue growth activity, activin/inhibin activity, | | |
| CC | chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, | | |
| CC | receptor/ligand activity, anti-inflammatory activity, cadherin/tumour | | |
| CC | invasion suppressor activity, and tumour inhibition activity. The DNAs | | |
| CC | are also stated to be useful for gene therapy. A host cell transfected | | |
| CC | with the DNA, or its subfragments and variants is useful for recombinant | | |
| CC | production of the human secreted protein clones. | | |
| XX | | | |
| SQ | Sequence 2930 BP; 902 A; 699 C; 692 G; 637 T; 0 other; | | |
| | Query Match 99.8%; Score 1024.4; DB 20; Length 2930; | | |
| | Best Local Similarity 99.9%; Pred. No. 7e-251; | | |
| | Matches 1025; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | | |
| Qy | 1 ATGCAACAAGATTGATGAGGTCGAGGTAGACACTGAAGTAGCCAACTTATCACTGATT 60 | | |
| | | | |

ID AAX27858 standard; DNA; 3685 BP.
XX AC AAX27858;
XX DT 02-JUN-1999 (first entry)
XX DE Human CSR3 protein coding sequence.
XX KW Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage;
KW scavenger receptor protein; intracellular stress; arteriosclerosis;
KW diabetic circulatory obstruction; microbial infection; ss.
XX OS Homo sapiens.
XX PN WO9909159-A1.
XX PD 25-FEB-1999.
XX PF 12-AUG-1998; 98WO-JP03602.
XX PR 30-JUL-1998; 98JP-0230121.
XX PR 13-AUG-1997; 97JP-0233396.
XX PA (NLSB) JAPAN TOBACCO INC.
XX PI Nakamura Y, Tokino T;
XX WPI; 1999-181032/15.
XX DR P-PSDB; AAY00994.
XX PT Scavenger receptor proteins - for treatment and diagnosis of
PT disorders involving cell stress
XX PS Claim 6; Page 142-150; 175pp; Japanese.
XX CC This sequence encodes the human cellular stress response 3 (CSR3) protein
CC of the invention. The CSR proteins are macrophage scavenger receptor
CC proteins. The CSR proteins can be used in the treatment, gene therapy
CC and diagnosis of diseases in which intracellular stress is important,
CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial
CC infection. Expression of the proteins is induced in vivo in response to
CC intracellular stress, and inhibits cell death as a result of such stress.
XX SQ Sequence 3685 BP; 901 A; 1142 C; 972 G; 669 T; 1 other;
Query Match 15.4%; Score 157.6; DB 20; Length 3685;
Best Local Similarity 55.5%; Pred. No. 5.3e-30;
Matches 304; Conservative 0; Mismatches 244; Indels 0; Gaps 0;
Qy 14 TGATGAGTCTGAGGTAGACACTGAAGTAGCCAACTTATCATGATGATTATGAAGAATGA 73
Db 1406 TGCTCAGTCCCGGCTGGACCTCAACGTCCGGACCTCTCCATGATCGTGGAGGATGA 1465
Qy 74 AGCTAGTAGACTCAAGCATGGTCAGTCATCAAGATTTTACATACTACAAGTCCAC 133
Db 1466 AGGCAGTGGACACACAGCATGGAGAAATCTTCGCAATGTCCATCTCCTAGAGGTGCC 1525
Qy 134 CGGGCCCCAGGGTCCAGAGAGTGCAGAGGATCCAGAGACCCCTCGGCCCAACTGGCA 193
Db 1526 CCGGCCCTCCAGGACCAAGAGGATTCAAAGGAGATATGGGCGTGAAGGGCTGTGTGGCG 1585
Qy 194 ACAAGGGACAGAAAGGAGAGAGGGGAGCCTGGACCACTTGGCCCTCGGGGTGAGAGAG 253
Db 1586 GCAGAGGCCCGAAGGAGACCCCGCATCTTTGGGCCCTCGGACCCAGGGTCTCAGG 1645
Qy 254 GCCCAATTGGACACAGCTGGTCCCGCCGAGAGCGTGGCGCAAGGATCTAAAGGCTCCC 313
Db 1646 GGCAACCTGGAGAGCGCGGCTGTGGAGAAAGGGGCCCTGTGGCCCTCAGGGTTCC 1705
Qy 314 AGGGCCCCAAGAGTCCCGTGTTCCTTGGGAAGCCCGGCCCTCAGGGCCCCAGTGGGG 373
Db 1706 CAGGCCCTCAAGAGGCTCAAGAGGCGACTTTGGAAGTGGAGGGCCGAGAGGACGCGCC 1765

Qy 374 ACCCAGGCCCGCGGGCCACACAGGCAAGAGGAGTCCCGGCCCTCAGGGCCCTCCTG 433
Db 1766 CAAAAGGGACATAGGCCCCCAGAGGCCCCCGGGGTCTCCAGGGCCCTCAG 1825
Qy 434 GCTTCCAGGACTTTCAGGGCACCGTTGGGGAGCCCTGGGGTGCCTGGACCTCGGGGACTGC 493
Db 1826 GGCTCAGGGAAACCGGGAATTGCAGGGAAGACAGGGTCCACAGGCCAGCGGGGGGCA 1885
Qy 494 CAGGCTTGCCTTGGGTACACAGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCCTGGGCC 553
Db 1886 TGGGGCCTAAGGTGAACACAGGGATCAGGGTCCCCCTGGTCTCCCGGGGCTCCAGGTC 1945
Qy 554 CATCAGGA 561
Db 1946 CACCAGGA 1953

Search completed: March 21, 2003, 02:19:27
Job time : 226.994 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:02:25 ; Search time 3431.41 Seconds
(without alignments)
11492.169 Million cell updates/sec

Title: US-09-763-712A-1_COPY_670_2024
Perfect score: 1355
Sequence: 1 atgcaacaagatttgatgag.....agaaaaaggttatcatcccg 1355

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.in.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|------------|---------------------|
| 1 | 1339 | 98.8 | 2983 | 9 | AB005145 | AB005145 Homo sapi |
| 2 | 1337.4 | 98.7 | 1886 | 9 | HS0802985 | AL713657 Homo sapi |
| 3 | 1337.4 | 98.7 | 2005 | 6 | AX454442 | AX454442 Sequence |
| 4 | 1337.4 | 98.7 | 2005 | 6 | AX490920 | AX490920 Sequence |
| 5 | 1337.4 | 98.7 | 2641 | 6 | AX047353 | AX047353 Sequence |
| 6 | 1337.4 | 98.7 | 3058 | 9 | AB038518 | AB038518 Homo sapi |
| 7 | 895.2 | 66.1 | 2637 | 10 | AB078434 | AB078434 Homo sapi |
| 8 | 887.4 | 65.5 | 3291 | 10 | AB038519 | AB038519 Mus muscu |
| 9 | 722.4 | 53.3 | 4330 | 9 | AB052103 | AB052103 Homo sapi |
| c 10 | 488.8 | 36.1 | 165088 | 2 | AC016128 | AC016128 Homo sapi |
| c 11 | 488.8 | 36.1 | 188439 | 9 | AP000915 | AP000915 Homo sapi |
| 12 | 484.6 | 35.8 | 71044 | 2 | AC024368 | AC024368 Homo sapi |
| 13 | 466.2 | 34.4 | 187635 | 2 | AP001022 | AP001022 Homo sapi |
| 14 | 465.8 | 34.4 | 178022 | 2 | AP000900 | AP000900 Homo sapi |
| c 15 | 437.4 | 32.3 | 188255 | 2 | AP000939 | AP000939 Homo sapi |
| 16 | 388 | 28.6 | 182029 | 2 | AC114677 | AC114677 Mus muscu |
| 17 | 388 | 28.6 | 193208 | 2 | AC102618 | AC102618 Mus muscu |
| 18 | 378.4 | 27.9 | 130763 | 2 | AC112416 | AC112416 Rattus no |
| c 19 | 271 | 20.0 | 178022 | 2 | AP000900 | AP000900 Homo sapi |
| c 20 | 249 | 18.4 | 415 | 6 | AX334493 | AX334493 Sequence |
| c 21 | 164.4 | 12.1 | 187635 | 2 | AP001022 | AP001022 Homo sapi |
| 22 | 157.6 | 11.6 | 3636 | 9 | AB007829 | AB007829 Homo sapi |
| 23 | 157.6 | 11.6 | 3685 | 6 | E32511 | E32511 Scavenger r |
| 24 | 157.6 | 11.6 | 3810 | 6 | E32509 | E32509 Scavenger r |
| 25 | 150.6 | 11.1 | 2215 | 10 | BC026446 | BC026446 Mus muscu |
| 26 | 129 | 9.5 | 6156 | 10 | AB009993 | AB009993 Mus muscu |
| 27 | 129 | 9.5 | 188255 | 2 | AP000939 | AP000939 Homo sapi |
| 28 | 128.2 | 9.5 | 2193 | 3 | PALCOLIA | M25282 P.lividus a |
| 29 | 125.8 | 9.3 | 6076 | 10 | AF1272661 | AF1272661 Rattus no |
| 30 | 124.2 | 9.2 | 1041 | 4 | AF138883 | AF138883 Bos tauru |
| 31 | 124.2 | 9.2 | 2053 | 5 | AB008374 | AB008374 Oncorhync |
| 32 | 124.2 | 9.2 | 4502 | 5 | AB052836 | AB052836 Oncorhync |
| 33 | 122.6 | 9.0 | 4995 | 10 | MMU16789 | U16789 Mus musculu |
| 34 | 121.6 | 9.0 | 810 | 6 | AR014116 | AR014116 Sequence |
| 35 | 121.4 | 9.0 | 4628 | 4 | AB008683 | AB008683 Bos tauru |
| 36 | 121.4 | 9.0 | 5676 | 6 | E07265 | E07265 cDNA encodi |
| 37 | 121.4 | 9.0 | 5676 | 9 | HUMCALV | D90279 Human mRNa |
| 38 | 121.4 | 9.0 | 7138 | 9 | HUMPALV | M76729 Human pro-a |
| 39 | 121 | 8.9 | 4581 | 4 | ECU62528 | U62528 Equus cabal |
| 40 | 119.8 | 8.8 | 2010 | 9 | HUMA2XICOL | L18987 Human alpha |
| 41 | 119.4 | 8.8 | 5551 | 10 | AF1272662 | AF1272662 Rattus no |
| 42 | 118.4 | 8.7 | 6109 | 6 | AX239611 | AX239611 Sequence |
| 43 | 118.4 | 8.7 | 6109 | 10 | AF176645 | AF176645 Mus muscu |
| 44 | 118 | 8.7 | 756 | 6 | AR014110 | AR014110 Sequence |
| 45 | 118 | 8.7 | 6114 | 10 | CRUPALV | M76730 Chinese ham |

ALIGNMENTS

RESULT 1
LOCUS AB005145
DEFINITION Homo sapiens CL-P1 mRNA for collectin placenta 1, complete cds.
ACCESSION AB005145
VERSION AB005145.1 GI:17026100
KEYWORDS Homo sapiens female tissue.lib:placenta cDNA to mRNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H.,
Sakai,Y., Fukuh,A., Sakamoto,T., Itabe,H., Suzutani,T.,
Ogasawara,M., Yoshida,I. and Wakamiya,N.

TITLE The membrane-type collectin CL-Pl is a scavenger receptor on
vascular endothelial cells
J. Biol. Chem. 276 (47), 44222-44228 (2001)
MEDLINE 21570232
REFERENCE 2 (bases 1 to 2983)
AUTHORS Ohtani,K.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1997) Katsuki Ohtani, Asahikawa Medical College,
Department of Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa,
Hokkaido 078-8510, Japan [E-mail: ohtani@asahikawa-med.ac.jp,
Tel: +81-166-68-2393, Fax: +81-166-68-2399]
FEATURES Location/Qualifiers

source
1..2983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="female"
/tissue_lib="placenta"
gene
1..2983
/gene="CL-Pl"
CDS
71..2299
/gene="CL-Pl"
/codon_start=1
/product="collectin placenta 1"
/protein_id="BAB72147.1"
/db_xref="GI:17026101"
/translation="MKDDFAEEEVQSFYKRFQIEQSTQCTCKNNWALKFSIILY
ILCALLITVAILGYKVYKMDNVTGMMSTSRQYTDKLTAVESDLKLQDQTKKAI
STNSLSFTRIDLDROQLREITEKTSKNKDTLEKQASGDALVDROSLKETLENN
SFLITVNTKTLQAVNGVYTNLQODTSVLQGNLQNMVSHNVYINNLNLTQVQORN
LITNLSQVSDTSQAIQRIKNDQNLQVFLQAKDDWLKQVSLQTLAANNSALA
KANNDLEDMSQNSFTQGMENITLISQNEQNLKLDQDLKDAENRTAKFNQLE
RFQLEFDIVNLIISNYSYTAHLRLTSLNNEVRTCTDLTKHTDDTLNNTLANI
RLDSVRMOQDLMRSLRDLTEVANLSVIMEEMKLVDSKHGQLIKNFTLQPPGPRGP
RGDRSGCPGPTNKGQKSGKGPDPGPPGKGLPGQPPGQGLQTVGPVGPVPLPG
GSRSPGPKPGDPGPPGKGLPGQPPGKGLPGQPPGQGLQTVGPVGPVPLPG
LPVGPVGPVPLPGQPPGKGLPGQPPGKGLPGQPPGQGLQTVGPVGPVPLPG
KEIFEDAKLFCEKSDSHLFIINTREEQWIKQWVRESHWIGLTDGERENENWLDG
TSPDYKNWKGQPNWGHGHPGEDCAGLIYAGWNDFQCEDVNFICEKDRITVLSS
AL"

BASE COUNT 914 a 707 c 703 g 659 t

Query Match 98.8%; Score 1339; DB 9; Length 2983;
Best Local Similarity 99.9%; Pred. No. 1.4e-300;
Matches 1350; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 ATGCAACAGATTGATGAGTTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
DB 1271 ATGCAACAGATTGATGAGTTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 1330
QY 61 ATGGAAGAAATGAAGCTAGTAGACTCCAGCATGCTGCTCATCAAGAATTTTACAATA 120
DB 1331 ATGGAAGAAATGAAGCTAGTAGACTCCAGCATGCTGCTCATCAAGAATTTTACAATA 1390
QY 121 CTACAGGTTCCACGGGCCCCAGGGTCCACAGAGGTGACAGAGGATCCACAGGACCCCGCT 180
DB 1391 CTACAGGTTCCACGGGCCCCAGGGTCCACAGAGGTGACAGAGGATCCACAGGACCCCGCT 1450
QY 181 GCGCCAACTGGCAACAGGACAGAAAGAGAGAGAGGGGAGCCCTGGACCACTGGCCCT 240
DB 1451 GCGCCAACTGGCAACAGGACAGAAAGAGAGAGAGGGGAGCCCTGGACCACTGGCCCT 1510
QY 241 GCGGGTGTAGAGAGGCGCAATTTGGACAGCTGCTGCCCGGAGAGCGTGGCGGCAAGGA 300
DB 1511 GCGGGTGTAGAGAGGCGCAATTTGGACAGCTGCTGCCCGGAGAGCGTGGCGGCAAGGA 1570
QY 301 TCTAAGGCTCCACGGGCCCCAAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGGCCCTCAG 360
DB 1571 TCTAAGGCTCCACGGGCCCCAAAGGCTCCCGTGGTTCCTCTGGGAAGCCCGGCCCTCAG 1630
QY 361 GCGCCCACTGGGACCCAGGCCCCCGGCCCAACAGGACAGAGAGGGACTCCCGGCCCT 420
DB 1631 GCGCCCACTGGGACCCAGGCCCCCGGCCCAACAGGACAGAGAGGGACTCCCGGCCCT 1690

QY 421 CAGGCCCTCTCGGCTTCCAGGACTTCCAGGCACCGTTGGGAGCCCTGGGGTGCTGGA 480
DB 1691 CAGGCCCTCTCGGCTTCCAGGACTTCCAGGCACCGTTGGGAGCCCTGGGGTGCTGGA 1750
QY 481 CCTCGGGACTGCCAGGCTTGCCTGGGTACCAGCATGCCAGGCCCCCAAGGCCCCCC 540
DB 1751 CCTCGGGACTGCCAGGCTTGCCTGGGTACCAGCATGCCAGGCCCCCAAGGCCCCCC 1810
QY 541 GSCCTCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCCAACCCG 600
DB 1811 GSCCTCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCCAACCCG 1870
QY 601 GCACCGGAGGACAATGGCTGCCCGCTCACTGGAAAGAACTTCCACAGACAAATGCTACTAT 660
DB 1871 GCACCGGAGGACAATGGCTGCCCGCTCACTGGAAAGAACTTCCACAGACAAATGCTACTAT 1930
QY 661 TTTTCAGTTGAGAAAATAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTCA 720
DB 1931 TTTTCAGTTGAGAAAATAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTCA 1990
QY 721 CATCTTGTTCATAAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGG 780
DB 1991 CATCTTGTTCATAAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGG 2050
QY 781 AGAGAGGCCACTGGATCGGCTCACAGACTCAGAGCGGTGAAATGAAATGAAATGAAATGAAATG 840
DB 2051 AGAGAGGCCACTGGATCGGCTCACAGACTCAGAGCGGTGAAATGAAATGAAATGAAATGAAATG 2110
QY 841 GATGGGACATCTCCAGACTACAAAATTTTGAAGCTGGACAGCCGGAATCACTGGGGTCAAT 900
DB 2111 GATGGGACATCTCCAGACTACAAAATTTTGAAGCTGGACAGCCGGAATCACTGGGGTCAAT 2170
QY 901 GCCCATGGCCAGGAGAGACTGTCTGGTGTGATTTATGCTGGGAGTGAACGATTTTC 960
DB 2171 GCCCATGGCCAGGAGAGACTGTCTGGTGTGATTTATGCTGGGAGTGAACGATTTTC 2230
QY 961 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAGACAGGAGACAGTACTGTCATCT 1020
DB 2231 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAGACAGGAGACAGTACTGTCATCT 2290
QY 1021 GCATTTATACGAGCTGTGATGGATCAGATGACAAATTTTTCAGCTCTCAAGAGCAAGG 1080
DB 2291 GCATTTATACGAGCTGTGATGGATCAGATGACAAATTTTTCAGCTCTCAAGAGCAAGG 2350
QY 1081 ACACCTCTTTTAAATTCATCACCCTTCTCATCAGATTGAAAAAAGAAAGCACTGAAA 1140
DB 2351 ACACCTCTTTTAAATTCATCACCCTTCTCATCAGATTGAAAAAAGAAAGCACTGAAA 2409
QY 1141 CCAATTACTGAAAAAATTTGACAGTGTGTTTTTACCATCCGTCATTTACCCAAAGAC 1200
DB 2410 CCAATTACTGAAAAAATTTGACAGTGTGTTTTTACCATCCGTCATTTACCCAAAGAC 2469
QY 1201 TTGGGAACATAAATGTTCCCGAGGATGATGCTGATTTTCATTTGTCACATGGAGTGA 1260
DB 2470 TTGGGAACATAAATGTTCCCGAGGATGATGCTGATTTTCATTTGTCACATGGAGTGA 2529
QY 1261 TCACATAGATTCTCTCCGTCAGTAACCGTGGGATTTATACAAATTTATGCTTCCAAAGTA 1320
DB 2530 TCACATAGATTCTCTCCGTCAGTAACCGTGGGATTTATACAAATTTATGCTTCCAAAGTA 2589
QY 1321 TGGAACTCTCCAAATCAGAAAAAGGTTATCAT 1351
DB 2590 TGGAACTCTCCAAATCAGAAAAAGGTTATCAT 2620

RESULT 2
HSM802985
LOCUS HSM802985 1886 bp mRNA linear PRI 20-MAR-2002
DEFINITION Homo sapiens mRNA; cDNA DKFp547G1215 (from clone DKFp547G1215).
ACCESSION AL713657
VERSION AL713657.1 GI:19584339
KEYWORDS

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1886)
AUTHORS Koehrer, K., Beyer, A., Mewes, H.W., Weil, B. and Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2002) MIPS, Am Klopfersplitz 18a, D-82152 Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; Sequenced by BMFZ (Biomedical Research Center at the Charité, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFzp547G1215) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/
FEATURES
source
1..1886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18p11.3"
/clone="DKFzp547G1215"
/tissue_type="brain"
/clone_lib="547 (synonym: hfbri). Vector pSport1; host DH10B; sites NotI + SalI"
/dev_stage="fetal"
1..1886
/gene="DKFzp547G1215"
/note="similarity to asialoglycoprotein receptor (Mus musculus)"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAD28466.1"
/db_xref="GI:19584340"
/translation="AIFKNLEERFOLFETDIVNIISVTAHHLRLTSLNLEVRT TCITDLNKHDDTLSTLANLIFLDSVLSLRMOODLMSRLDTEVANLSVIMEKLV DSKHQLIKFTILQGPGRGDRGSGQPGPTGNKGQKGEKGPGRGPRGERG PIGPAGPGRGGRGSGKSGPGKPGQSGSDPGPPGPGKGLPQPGP PFGLOGTGVGPGVGRGLPGLPGVPGMPKPGPPGPGVPLALONPTPA PEDNGCPHWNKFTDKCYFSVEKEIFEDAKLFCEKSKSHLVFNTRREQWIKQWV GRESHWITGLTDSRENEWKWLDTGSPDYKNWKAQOPDNWGHGHPGEDCAGLIYAGW NDFOCEDVNNFICEKRETVLSSAL"
polyA_signal 1838..1843
/gene="DKFzp547G1215"
polyA_site 1869
/gene="DKFzp547G1215"
BASE COUNT 569 a 444 c 445 g 428 t
ORIGIN
Query Match 98.7%; Score 1337.4; DB 9; Length 1886;
Best Local Similarity 99.9%; Pred. No. 3.2e-300;
Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 1 ATGCAACAAGATTTCATCAGGTTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
Db 228 ATGCAACAAGATTTCATCAGGTTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 287
Qy 61 ATGGAAGAAATGAAGCTAGTAGACTCCCAAGCATGTCAGCTCATCAAGAATTTTACAATA 120
Db 288 ATGGAAGAAATGAAGCTAGTAGACTCCCAAGCATGTCAGCTCATCAAGAATTTTACAATA 347
Qy 121 CTACAAGTCCACCGGGCCCGACAGGGGTCCCAAGAGGTGACAGAGATCCACAGGACCCCT 180
Db 348 CTACAAGTCCACCGGGCCCGACAGGGGTCCCAAGAGGTGACAGAGATCCACAGGACCCCT 407
Qy 181 GGCCCACTGTCACACAGGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 408 GGCCCACTGTCACACAGGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 467

Qy 241 GCGGTGAGAGAGGCCCCAATTGGACAGCTGGTCCCCCGGAGAGCGTGGCGGCAAGGA 300
Db 468 GCGGTGAGAGAGGCCCCAATTGGACAGCTGGTCCCCCGGAGAGCGTGGCGGCAAGGA 527
Qy 301 TCTAAAGGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCCGGCTCAG 360
Db 528 TCTAAAGGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCCGGCTCAG 587
Qy 361 GGCCCACTGGGGACCCAGGCCCCCGGCCCCACAGCAAAAGAGGACTCCCGGCGCT 420
Db 588 GGCTCAGTGGGGACCCAGGCCCCCGGCCCCACAGCAAAAGAGGACTCCCGGCGCT 647
Qy 421 CAGGCCCTCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCTGGGTGCTGGA 480
Db 648 CAGGCCCTCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCTGGGTGCTGGA 707
Qy 481 CCTCGGGGACTGCCAGGCTTGCTGGGTACAGGCATGCCAGGCCCAAGGGCCCCCCC 540
Db 708 CCTCGGGGACTGCCAGGCTTGCTGGGTACAGGCATGCCAGGCCCAAGGGCCCCCCC 767
Qy 541 GSCCTCTCTGCCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAAGAGCAACCCG 600
Db 768 GSCCTCTCTGCCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAAGAGCAACCCG 827
Qy 601 GCACCGGAGGACAATGGCTGCCCGCTCCTCAGTGAAGAACTTCACAGACAAATGCTACTAT 660
Db 828 GCACCGGAGGACAATGGCTGCCCGCTCCTCAGTGAAGAACTTCACAGACAAATGCTACTAT 887
Qy 661 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAGAGCAAGTCTTCA 720
Db 888 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAGAGCAAGTCTTCA 947
Qy 721 CATCTTTGTTTCATAAACACATAGAGAGCAAGCAATGGATAAAAGACAGATGTTAGG 780
Db 948 CATCTTTGTTTCATAAACACATAGAGAGCAAGCAATGGATAAAAGACAGATGTTAGG 1007
Qy 781 AGAGAGAGCCACTGGATCGGCTCAGACTCAGAGCTGAGAGCTGAAATGAATGGAAGTGGCTG 840
Db 1008 AGAGAGAGCCACTGGATCGGCTCAGACTCAGAGCTGAGAGCTGAAATGAATGGAAGTGGCTG 1067
Qy 841 GATGGGACATCTCCAGACTACAAAAATTTGGAAGCTGGACAGCCCGGATTAATCGGGTCA 900
Db 1068 GATGGGACATCTCCAGACTACAAAAATTTGGAAGCTGGACAGCCCGGATTAATCGGGTCA 1127
Qy 901 GSCCATGGCCAGGAGAGAGCTGCTGGTGTGATTATGCTGGGCGAGTGAACGATTTTC 960
Db 1128 GSCCATGGCCAGGAGAGAGCTGCTGGTGTGATTATGCTGGGCGAGTGAACGATTTTC 1187
Qy 961 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAAGACAGGAGACAGTACTGTCATCT 1020
Db 1188 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAAGACAGGAGACAGTACTGTCATCT 1247
Qy 1021 GCATTTATACGAGCTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAGGCAAGAG 1080
Db 1248 GCATTTATACGAGCTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAGGCAAGAG 1307
Qy 1081 ACACCTCTTTCTTAATTGATCACTTCTCATCAGATTGAAAAAAGACACTGAAAA 1140
Db 1308 ACACCTCTTTCTTAATTGATCACTTCTCATCAGATTG-AAAAAAGACACTGAAAA 1366
Qy 1141 CCAATTACTGAAAAAATTTGACAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC 1200
Db 1367 CCAATTACTGAAAAAATTTGACAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC 1426
Qy 1201 TTGGGAACATAAATGTTCCCGAGGGTGATATGCTGATTTTCATTTGTCACATGGACTGAA 1260
Db 1427 TTGGGAACATAAATGTTCCCGAGGGTGATATGCTGATTTTCATTTGTCACATGGACTGAA 1486
Qy 1261 TCACATAGATTCTCTCTCGTCAAGTAAACCGTGCAGTATATACAAATTTATGTCCTCCAAAGTA 1320
Db 1487 TCACATAGATTCTCTCTCGTCAAGTAAACCGTGCAGTATATACAAATTTATGTCCTCCAAAGTA 1546

QY 1321 TGAACACTCCAATCAGAAAAAGTTATCAT 1351
Db 1547 TGAACACTCCAATCAGAAAAAGTTATCAT 1577

RESULT 3
AX454442
LOCUS AX454442 2005 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 27 from Patent WO0208284.
ACCESSION AX454442
VERSION AX454442.1 GI:21713845
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
AUTHORS Baker,K.P., Ferrara,N., Gerber,H., Gerritsen,M.E., Goddard,A.,
Godowski,P.J., Gurney,A.L., Hillan,K.J., Marsters,S.A., Pan,J.,
Paoni,N.F., Stephan,J.P., Watanabe,C.K., Williams,P.M., Wood,W.I.
and Ye,W.
TITLE Compositions and methods for the diagnosis and treatment of
disorders involving angiogenesis
JOURNAL Patent: WO 0208284-A 27 31-JAN-2002;
Genentech, Inc. (US) ; Baker, Kevin P. (US) ; Ferrara, Napoleone
(US) ; Gerber, Hanspeter (US) ; Gerritsen, Mary E. (US) ; Goddard,
Audrey (US) ; Godowski, Paul J. (US) ; Gurney, Austin L. (US) ;
Hillan, Kenneth J. (US) ; Marsters, Scot A. (US) ; Pan, James (US)
; Paoni, Nicholas F. (US) ; Stephan, Jean-Philippe F. (US) ;
Watanabe, Colin K. (US) ; Williams, P. Mickey (US) ; Wood, William
I. (US)

FEATURES
source Location/Qualifiers
1..2005
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 606 a 493 c 491 g 415 t
ORIGIN

Query Match 98.7%; Score 1337.4; DB 6; Length 2005;
Best Local Similarity 99.9%; Pred. No. 3.3e-300;
Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGCAACAAGATTGATAGGTCGAGGTAGACACTGAAGTAGCCAACTTATCAGTGATT 60
Db 478 ATGCAACAAGATTGATAGGTCGAGGTAGACACTGAAGTAGCCAACTTATCAGTGATT 537
QY 61 ATGGGAAGAAATGAAGTAGTAGACTCCAGCATGCTCAGCTCATCAAGAATTTTACAATA 120
Db 538 ATGGGAAGAAATGAAGTAGTAGACTCCAGCATGCTCAGCTCATCAAGAATTTTACAATA 597
QY 121 CTACAAGGTCCACCGGGCCCCAGGGTCCAAGAGGTGACAGAGGATCCCGAGGACCCCT 180
Db 598 CTACAAGGTCCACCGGGCCCCAGGGTCCAAGAGGTGACAGAGGATCCCGAGGACCCCT 657
QY 181 GCGCCACTGGCAACAAGGGACAGAAAGAGAGAGAGGGGAGCCCTGGACCACCTGGCCCT 240
Db 658 GCGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGAGGGGAGCCCTGGACCACCTGGCCCT 717
QY 241 GCGGGTGTAGAGAGGCCAATTTGACACAGCTGCTCCCGCCGGAGAGCGTGGCGGCAAGGA 300
Db 718 GCGGGTGTAGAGAGGCCAATTTGACACAGCTGCTCCCGCCGGAGAGCGTGGCGGCAAGGA 777
QY 301 TCTAAGAGGTCCACGGGGCCCCAAAGGCTCCCGTGGTTCCCTTGGGAAGCCCGGCCCTCAG 360
Db 778 TCTAAGAGGTCCACGGGGCCCCAAAGGCTCCCGTGGTTCCCTTGGGAAGCCCGGCCCTCAG 837
QY 361 GCGCCAGTGGGAGCCAGGCCCCCGGGCCCCACAGGCAAGAGAGGACTTCCCGGCCCT 420
Db 838 GCGCCAGTGGGAGCCAGGCCCCCGGGCCCCACAGGCAAGAGAGGAGTCCCGGCCCT 897
QY 421 CAGGCGCTCTCTGGCTTCCAGGAGCTTCAGGGCACCGTTGGGGAGCCTGGGGTGCTGGA 480
Db 898 CAGGCGCTCTCTGGCTTCCAGGAGCTTCAGGGCACCGTTGGGGAGCCTGGGGTGCTGGA 957

QY 481 CCTCGGGAGACTGCCAGGCTTGCTTGGGTACCAGGATGCCAGGCCCGCCCAAGGGCCCCCCC 540
Db 958 CCTCGGGAGACTGCCAGGCTTGCTTGGGTACCAGGATGCCAGGCCCGCCCAAGGGCCCCCCC 1017
QY 541 GGCCTCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTCAGNAATGAGCAACCCCG 600
Db 1018 GGCCTCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTCAGNAATGAGCAACCCCG 1077
QY 601 GCACCGGAGGACAATGGGTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 660
Db 1078 GCACCGGAGGACAATGAGTGGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 1137
QY 661 TTTTCAGTTGAGAAAAAATTTTTCAGGATGCAAAAGCTTTTCTGTGCAAGCAAGCTTCA 720
Db 1138 TTTTCAGTTGAGAAAAAATTTTTCAGGATGCAAAAGCTTTTCTGTGCAAGCAAGCTTCA 1197
QY 721 CATCTTGTTCATAAACAACACTAGAGAGGAGNACAGCAATGGATAAAAAACAGATGCTAGG 780
Db 1198 CATCTTGTTCATAAACAACACTAGAGAGGAGNACAGCAATGGATAAAAAACAGATGCTAGG 1257
QY 781 AGAGAGAGCAGCTGGATCGGCCTCACAGACTCAGAGCGTGAAAAATGAATGGAAGTGGCTG 840
Db 1258 AGAGAGAGCAGCTGGATCGGCCTCACAGACTCAGAGCGTGAAAAATGAATGGAAGTGGCTG 1317
QY 841 GATGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAAT 900
Db 1318 GATGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAAT 1377
QY 901 GGCATGGCCAGGAGAGAGACTGTGCTGGTGTGATTTATGCTGGCAGTGGAAAGATTTC 960
Db 1378 GGCATGGCCAGGAGAGAGACTGTGCTGGTGTGATTTATGCTGGCAGTGGAAAGATTTC 1437
QY 961 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAAGACAGGAGACAGTACTGTCTATCT 1020
Db 1438 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAAGACAGGAGACAGTACTGTCTATCT 1497
QY 1021 GCATTTAATACGGACTGTGATGGATCAGATGAGCAAAATTTTTCAGCTCTCAAAGGCAAGG 1080
Db 1498 GCATTTAATACGGACTGTGATGGATCAGATGAGCAAAATTTTTCAGCTCTCAAAGGCAAGG 1557
QY 1081 ACATCTCTTTCTAATTCATCACCCTTCATCAGATGGAAGAAAAAGACACTGAAGA 1140
Db 1558 ACATCTCTTTCTAATTCATCACCCTTCATCAGATGGAAGAAAAAGACACTGAAGA 1616
QY 1141 CCAATTACTGAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC 1200
Db 1617 CCAATTACTGAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCATTACCCAAAGAC 1676
QY 1201 TTGGGAACATAAATGTTCCCGAGGTGATATGCTGATTTTCATTGTGACATGGAGTAA 1260
Db 1677 TTGGGAACATAAATGTTCCCGAGGTGATATGCTGATTTTCATTGTGACATGGAGTAA 1736
QY 1261 TCACATAGATTCCTCCGTCAGTAACCGTGGGATATACAATTTATGCTTTCCAAAGTA 1320
Db 1737 TCACATAGATTCCTCCGTCAGTAACCGTGGGATATACAATTTATGCTTTCCAAAGTA 1796
QY 1321 TGAACACTCCAATCAGAAAAAGTTATCAT 1351
Db 1797 TGAACACTCCAATCAGAAAAAGTTATCAT 1827

RESULT 4
AX490920
LOCUS AX490920 2005 bp DNA linear PAT 16-AUG-2002
DEFINITION Sequence 27 from Patent WO0200690.
ACCESSION AX490920
VERSION AX490920.1 GI:22323797
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1
AUTHORS Baker,K.P., Ferrara,N., Gerber,H., Gerritsen,M.E., Goddard,A.,
Godowski,P.J., Gurney,A.L., Hillan,K.J., Marsters,S.A., Pan,J.,
Paoni,N.F., Stephan,J.P., Watanabe,C.K., Williams,P.M., Wood,W.I.
and Ye,W.
TITLE Compositions and methods for the diagnosis and treatment of
disorders involving angiogenesis
JOURNAL Patent: WO 0200690-A 27 03-JAN-2002;
Genentech, Inc. (US)
FEATURES Location/Qualifiers
Source 1..2005
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 606 a 493 c 491 g 415 t
ORIGIN
Query Match 98.7%; Score 1337.4; DB 6; Length 2005;
Best Local Similarity 99.9%; Pred. No. 3.3e-300;
Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 1 ATGCAACAAAGATTGATGAGGTTCGAGGTAGACACTGAAGTAGCCAACTTATFCAGTGATT 60
Db 478 ATGCAACAAAGATTGATGAGGTTCGAGGTAGACACTGAAGTAGCCAACTTATFCAGTGATT 537
Qy 61 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAATTTTACAATA 120
Db 538 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAATTTTACAATA 597
Qy 121 CTACAAGTCCACGGGCCCCAGGGTCCAAAGGTGACAGGTGACAGAGTATCCAGGGACCCCT 180
Db 598 CTACAAGTCCACGGGCCCCAGGGTCCAAAGGTGACAGGTGACAGAGTATCCAGGGACCCCT 657
Qy 181 GGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGGGGGAGCCCTGGACCACTGGCCCT 240
Db 658 GGCCCAACTGGCAACAAGGGACAGAAAGAGAGAGGGGGAGCCCTGGACCACTGGCCCT 717
Qy 241 GCGGTGAGAGAGGCCCAATTTGGACCACTGGTCCCGCCGAGAGCGTGGGGGCAAAAGGA 300
Db 718 GCGGTGAGAGAGGCCCAATTTGGACCACTGGTCCCGCCGAGAGCGTGGGGGCAAAAGGA 777
Qy 301 TCTAAGGCTCCAGGGCCCCCAAGGCTCCGCTGGTGGTCCCTGGCAAGCCGCGCCCTCAG 360
Db 778 TCTAAGGCTCCAGGGCCCCCAAGGCTCCGCTGGTGGTCCCTGGCAAGCCGCGCCCTCAG 837
Qy 361 GGCCCCAGTGGGGACCCAGGCCCCCGGGCCACACAGGCAAGAGGGACTCCCGCGCCCT 420
Db 838 GGCCCCAGTGGGGACCCAGGCCCCCGGGCCACACAGGCAAGAGGGACTCCCGCGCCCT 897
Qy 421 CAGGCCCTCTCTGGCTTCCAGGGACTTCAGGGCACCCTGGGGAGCCCTGGGGTGCTGGA 480
Db 898 CAGGCCCTCTCTGGCTTCCAGGGACTTCAGGGCACCCTGGGGAGCCCTGGGGTGCTGGA 957
Qy 481 CCTCGGGAGCTGCCAGGCTTCGCTGGGTACAGGCATGCCAGGCCCAAGGGCCCCCCCC 540
Db 958 CCTCGGGAGCTGCCAGGCTTCGCTGGGTACAGGCATGCCAGGCCCAAGGGCCCCCCCC 1017
Qy 541 GGCCTCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCCG 600
Db 1018 GGCCTCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCCG 1077
Qy 601 GCACCGGAGCAATGGCTGCCCGCTCAGTGGAAAGAACTTCACAGCAAAATGCTACTAT 660
Db 1078 GCACCGGAGCAATGAGCTGCGCGCTCAGTGGAAAGAACTTCACAGCAAAATGCTACTAT 1137
Qy 661 TTTTCAGTTGAAAGAAATTTTTCAGGATCAAGCTTTTCTGTGGAAGACAAGTCTTCA 720
Db 1138 TTTTCAGTTGAAAGAAATTTTTCAGGATCAAGCTTTTCTGTGGAAGACAAGTCTTCA 1197
Qy 721 CATCTGTTTTCATAAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGTTAGGG 780
Db 1198 CATCTGTTTTCATAAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGTTAGGG 1257
Qy 781 AGAGAGAGCCACTGGATCGGCTTCACAGACTCAGAGCGTGAAGATGAATGGAAGTGGCTG 840

Db 1258 AGAGAGAGCCACTGATCGGCTCTCAGACTCAGAGCGTGAAATCAATGGAAGTGGCTG 1317
Qy 841 GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCCGGATAACTGGGGTCAT 900
Db 1318 GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCCGGATAACTGGGGTCAT 1377
Qy 901 GCCATGGCCAGGAGAAAGACTGTCTGGTGTGATTATGCTGGCAGTGGAAACGATTC 960
Db 1378 GCCATGGCCAGGAGAAAGACTGTCTGGTGTGATTATGCTGGCAGTGGAAACGATTC 1437
Qy 961 CAATGTGAAGAGCTCAATAACTTTTCGCGAAAAGACAGGAGACAGTACTGTCATCT 1020
Db 1438 CAATGTGAAGAGCTCAATAACTTTTCGCGAAAAGACAGGAGACAGTACTGTCATCT 1497
Qy 1021 GCATTATACGAGCTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAAAGGCAAGG 1080
Db 1498 GCATTATACGAGCTGTGATGGGATCAGATGAGCAAAATTTTCAGCTCTCAAAAGGCAAGG 1557
Qy 1081 ACACCTCCTTTCTAATTGATCACCTTCTCATCAGATTGAAAAAAAAGACACTGAAAA 1140
Db 1558 ACACCTCCTTTCTAATTGATCACCTTCTCATCAGATTG-AAAAAAAAGACACTGAAAA 1616
Qy 1141 CCAATTACTGAAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCATATCCCAAAGAC 1200
Db 1617 CCAATTACTGAAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCATATCCCAAAGAC 1676
Qy 1201 TTGGAACATAAATCTTCCCGAGGTGATGCTGATTTTCATTTTTCGACATGGACTGAA 1260
Db 1677 TTGGAACATAAATCTTCCCGAGGTGATGCTGATTTTCATTTTTCGACATGGACTGAA 1736
Qy 1261 TCACATAGATTCTCTCCGTCAGTAACCGTCGATTATACAAATATGCTTCCCAAAGTA 1320
Db 1737 TCACATAGATTCTCTCCGTCAGTAACCGTCGATTATACAAATATGCTTCCCAAAGTA 1796
Qy 1321 TGGAACTCCCAATCAGAAAAAGGTTATCAT 1351
Db 1797 TGGAACTCCCAATCAGAAAAAGGTTATCAT 1827
RESULT 5
AX047353 AX047353 2641 bp DNA linear PAT 15-DEC-2000
LOCUS Sequence 39 from Patent WO0068380.
DEFINITION AX047353
ACCESSION AX047353
VERSION AX047353.1 GI:11876599
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 2641)
Bandman,O., Hillman,J.L., Tang,Y.T., Lal,P., Yue,H., Baughn,M.R.,
Lu,D.A. and Azimzai,Y.
TITLE Extracellular matrix and adhesion-associated proteins
JOURNAL Patent: WO 0068380-A 39 16-NOV-2000;
Incyte Genomics, Inc. (US)
FEATURES Location/Qualifiers
Source 1..2641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 4586187CBI"
BASE COUNT 823 a 628 c 608 g 582 t
ORIGIN
Query Match 98.7%; Score 1337.4; DB 6; Length 2641;
Best Local Similarity 99.9%; Pred. No. 3.3e-300;
Matches 1349; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 1 ATGCAACAAAGATTGATGAGGTTCGAGGTAGACACTGAAGTAGCCAACTTATFCAGTGATT 60
Db 912 ATGCAACAAAGATTGATGAGGTTCGAGGTAGACACTGAAGTAGCCAACTTATFCAGTGATT 971

QY 61 ATGGAAGAAATGAAGCTAGTAGCTCCCAAGCATGTGTACGTCTCATCAAGAATTTTCAATA 120
Db 972 ATGGAAGAAATGAAGCTAGTAGCTCCCAAGCATGTGTACGTCTCATCAAGAATTTTCAATA 1031
QY 121 CTACAGGTCCACCGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCCAAGGACCCCT 180
Db 1032 CTACAGGTCCACCGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCCAAGGACCCCT 1091
QY 181 GGCCCAACTGGCAACAAGGACAGAAAAGGAGAAAGGGGAGCCCTGGACACCTGGCCCT 240
Db 1092 GGCCCAACTGGCAACAAGGACAGAAAAGGAGAAAGGGGAGCCCTGGACACCTGGCCCT 1151
QY 241 GCGGTGAGAGAGGCCAATTTGGACCAAGTGTCTCCCGGAGAGCGTGGCGGCAAGGA 300
Db 1152 GCGGTGAGAGAGGCCAATTTGGACCAAGTGTCTCCCGGAGAGCGTGGCGGCAAGGA 1211
QY 301 TCTAAGGCTCCAGGGCCCCAAAGGCTCCGCTGGTTCCTCTGGGAAGCCCGGCCCTCAG 360
Db 1212 TCTAAGGCTCCAGGGCCCCAAAGGCTCCGCTGGTTCCTCTGGGAAGCCCGGCCCTCAG 1271
QY 361 GGCCCCAGTGGGACCCAGGCCCCCCGGGCCACCAGGCAAGAGGGACTCCCGGCCCT 420
Db 1272 GGCCCCAGTGGGACCCAGGCCCCCCGGGCCACCAGGCAAGAGGGACTCCCGGCCCT 1331
QY 421 CAGGGCCCTCTGGCTTCAGGGACTTCAGGGACCCGTGTGGGGAGCCCTGGGTGCTTGA 480
Db 1332 CAGGGCCCTCTGGCTTCAGGGACTTCAGGGACCCGTGTGGGGAGCCCTGGGTGCTTGA 1391
QY 481 CCTCGGGACTCCAGGCTTGCTGGGTACCAGGATGCCAGGCCCCCAAGGGCCCCCCC 540
Db 1392 CCTCGGGACTCCAGGCTTGCTGGGTACCAGGATGCCAGGCCCCCAAGGGCCCCCCC 1451
QY 541 GGCCCTCTCTGGCCCTCAGGAGCGTGTGTCCTGGCCCTGCAGAAAGGCAACCCCG 600
Db 1452 GGCCCTCTCTGGCCCTCAGGAGCGTGTGTCCTGGCCCTGCAGAAAGGCAACCCCG 1511
QY 601 GCACGGAGGACAATGGCTGCGCCCTCAGTGGAAAGCTTCACAGACAAATGCTACTAT 660
Db 1512 GCACGGAGGACAATGGCTGCGCCCTCAGTGGAAAGCTTCACAGACAAATGCTACTAT 1571
QY 661 TTTTCAGTTGAGAAAGAAATTTTCAGAGTCAAGAGCTTTCTGTGAAGACAGTCTCA 720
Db 1572 TTTTCAGTTGAGAAAGAAATTTTCAGAGTCAAGAGCTTTCTGTGAAGACAGTCTCA 1631
QY 721 CATCTGTTTTCATAAAGCTAGAGGAACCAATGGATAAAAAACAGATGGTAGG 780
Db 1632 CATCTGTTTTCATAAAGCTAGAGGAACCAATGGATAAAAAACAGATGGTAGG 1691
QY 781 AGAGAGCCACTGGATGGCCCTCAGACTCAGAGCGTGAATGAATGGAAGTGGCTG 840
Db 1692 AGAGAGCCACTGGATGGCCCTCAGACTCAGAGCGTGAATGAATGGAAGTGGCTG 1751
QY 841 GATGGACATCTCCAGACTACAAAAATTTGGAAGCTGGACAGCCGGAATCACTGGGTCTAT 900
Db 1752 GATGGACATCTCCAGACTACAAAAATTTGGAAGCTGGACAGCCGGAATCACTGGGTCTAT 1811
QY 901 GCGCATGGGCCAGGAGACTGTCTGGTGTGATTATGCTGGGAGTGAACGATTTTC 960
Db 1812 GCGCATGGGCCAGGAGACTGTCTGGTGTGATTATGCTGGGAGTGAACGATTTTC 1871
QY 961 CAATGTGAAGAGCTCAATAACTTCATTTGCGAAAAGACAGGAGACAGTACTGTCTAT 1020
Db 1872 CAATGTGAAGAGCTCAATAACTTCATTTGCGAAAAGACAGGAGACAGTACTGTCTAT 1931
QY 1021 GCATTATACGAGCTGTGATGGATCAGATGACAAATTTTCAGCTCTCAAGGCAAGG 1080
Db 1932 GCATTATACGAGCTGTGATGGATCAGATGACAAATTTTCAGCTCTCAAGGCAAGG 1991
QY 1081 ACACCTCTTTCTAATTTGATCATCACTTCTCATCAGATTTGAAAAAAGACACTGAAAA 1140
Db 1992 ACACCTCTTTCTAATTTGATCATCACTTCTCATCAGATTG- AAAAAAAGACACTGAAA 2050
QY 1141 CCAATTACTGAAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCTATTACCCCAAGAC 1200

Db 2051 CCAATTACTGAAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCGTCTATTACCCAAGAC 2110
QY 1201 TTGGGAACATAAATCTCCCAAGGTGATATGCTGATTTTCATTTCGACATGGACTGAA 1260
Db 2111 TTGGGAACATAAATCTCCCAAGGTGATATGCTGATTTTCATTTCGACATGGACTGAA 2170
QY 1261 TCACATAGATTCTCCTCCGTCTAGTAACCGTCGATTATACAAATATGCTTCCCAAGTA 1320
Db 2171 TCACATAGATTCTCCTCCGTCTAGTAACCGTCGATTATACAAATATGCTTCCCAAGTA 2230
QY 1321 TGGACACTCCAATCAGAAAAAGGTTATCAT 1351
Db 2231 TGGACACTCCAATCAGAAAAAGGTTATCAT 2261
RESULT 6
AB038518 3058 bp mRNA linear PRI 08-MAR-2001
DEFINITION Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
type I, complete cds.
ACCESSION AB038518
VERSION AB038518.1 GI:13365514
KEYWORDS
SOURCE Homo sapiens tissue_lib:Placenta cdna to mRNA.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Nakamura,K., Funakoshi,H., Miyamoto,K., Tokunaga,F. and Nakamura,T.
TITLE Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRCL), a novel member of a
scavenger receptor family
JOURNAL Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
MEDLINE 21092718
AUTHORS Nakamura,K. and Nakamura,T.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical
School, Division of Biochemistry, Biomedical Research Center; 2-2
Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail:knakamura@onbich.med.osaka-u.ac.jp,
Tel:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)
FEATURES
source Location/Qualifiers
1..3058
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_lib="Placenta"
1..3058
/gene="SRCL"
209..2437
/gene="SRCL"
/codon_start=1
/product="scavenger receptor with C-type lectin type I"
/protein_id="BAB39147.1"
/db_xref="gi:13365515"
/translation="MKDDFAEEVEVQSGYKRFQIEGCTCKKNWALFKFSILLY
ILCALLITVAILGVKVVKMDNVQSGMETSRQYDDKLTAVEDSKLKLQDQTKGKAI
STNSLSTFRSDILDRLQOLREITEKTSKNDKLEKLOASGDALVDROSLKETLENN
SFLITTVNKTLOAYNGVYNLQDTSVLOGNLQNMYSNVVIMNLLNLTQVQORN
LITNLQSVDDTSQAQRIKNDQNLQVFLQAKKDDTLWKEKVSQDGLTAANNALSA
KANNDLEDMSNLSFTGOMENTTTSQANEQNKLQDLQHLKDAENRTAKFNQLEE
RFQLFETDIVNIISIVTAHLRLTLTSQANEVRTCTDTLTKHTDDTLNNLTANI
RDSVSLRMOODLMKSLRDLTVANLSVIMEEMKLVDSKHGOLIKNFTLQCGPGRGP
RGDRSGPGPTGKQKGEKPGEPGAGERPIGPGAPGGERGKSGSGQPK
GSRGSPKPGQGGSPGPPGPGKEGLPGQPGPPGPGQGLQCTVPGGPGPGLPG
LPVPGMPGPGPPGPGSVAVFLAQNEPTPAPENSCPPHWKNTDKCIYFSVE
KEIFEDAKLFCEDKSSHLVFINTEEQWIKKQMWGRESHWIGLTDSSRENEWKLDG
TSPDYKNWKAQPDNMGHGGPGEDCAGLIYAGQWDFQEDVNNF ICEKORETVLSS
AL"
polyA_signal 3020..3025
/gene="SRCL"
BASE COUNT 911 a 764 g 639 t

ORIGIN

| | | | | | |
|-----------------------|------|--|---------------------|-----------|--------------|
| Query Match | | 98.7%; | Score 1337.4; | DB 9; | Length 3058; |
| Best Local Similarity | | 99.9%; | Pred. No. 3.4e-300; | | |
| Matches 1349; | | Conservative 0; | Mismatches 1; | Indels 1; | Gaps 1; |
| Qy | 1 | ATGCAACAAGATTGATGAGTGCAGGTTAGACACTGAAGTAGCCAACCTTATCAGTGATT | 60 | | |
| Db | 1409 | ATGCAACAAGATTGATGAGTGCAGGTTAGACACTGAAGTAGCCAACCTTATCAGTGATT | 1468 | | |
| Qy | 61 | ATGGAAGAATGAGCTAGTAGACTCCAAGCATGGTCAGTCATCAAGAATTTTACATA | 120 | | |
| Db | 1469 | ATGGAAGAATGAGCTAGTAGACTCCAAGCATGGTCAGTCATCAAGAATTTTACATA | 1528 | | |
| Qy | 121 | CTACAAGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGTCCACAGGAGCCGCCCT | 180 | | |
| Db | 1529 | CTACAAGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGTCCACAGGAGCCGCCCT | 1588 | | |
| Qy | 181 | GGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGAGGGGGAGCCTGGACACCTTGGCCCT | 240 | | |
| Db | 1589 | GGCCCAACTGGCAACAAGGGACAGAAAGGAGAGAGGGGGAGCCTGGACACCTTGGCCCT | 1648 | | |
| Qy | 241 | CGCGGTGAGAGAGGCCCAATTTGGACCAGCTGGTCCCCCGGAGAGCGTGGGGCAAGGA | 300 | | |
| Db | 1649 | CGCGGTGAGAGAGGCCCAATTTGGACCAGCTGGTCCCCCGGAGAGCGTGGGGCAAGGA | 1708 | | |
| Qy | 301 | TCTAAAGGCTCCACAGGGCCCCAAGGGTCCCGTGGTTCCTCCCTGGGAAGCCGCGCCCTCAG | 360 | | |
| Db | 1709 | TCTAAAGGCTCCACAGGGCCCCAAGGGTCCCGTGGTTCCTCCCTGGGAAGCCGCGCCCTCAG | 1768 | | |
| Qy | 361 | GGCCCCAGTGGGGACCAAGGCCCCCGGGGCCACAGCAAGAGGAGACTCCCGCGGCCCT | 420 | | |
| Db | 1769 | GGCCCCAGTGGGGACCAAGGCCCCCGGGGCCACAGCAAGAGGAGACTCCCGCGGCCCT | 1828 | | |
| Qy | 421 | CAGGSCCTCTGCTGCTCCAGGACTTTCAGGACCGCTTGGGAGCCCTGGGGTGCCTGGA | 480 | | |
| Db | 1829 | CAGGSCCTCTGCTGCTCCAGGACTTTCAGGACCGCTTGGGAGCCCTGGGGTGCCTGGA | 1888 | | |
| Qy | 481 | CCTCGGGGACTGCCAGGCTTGCCTGGGTACAGGCGATGCCAGGCCCAAGGGCCCCCCC | 540 | | |
| Db | 1889 | CCTCGGGGACTGCCAGGCTTGCCTGGGTACAGGCGATGCCAGGCCCAAGGGCCCCCCC | 1948 | | |
| Qy | 541 | GGCCCTCTGCCCCATCAGAGCGGTGGTGCCTTGGCCCTGCAGAAAGAGCCCAACCCCG | 600 | | |
| Db | 1949 | GGCCCTCTGCCCCATCAGAGCGGTGGTGCCTTGGCCCTGCAGAAAGAGCCCAACCCCG | 2008 | | |
| Qy | 601 | GCACGGAGGACATGGTGCCTGCCCTCAGTGGAGACTTTCAGACAAATGCTACTAT | 660 | | |
| Db | 2009 | GCACGGAGGACATAGCTGCCCGCTCAGTGGAGAACTTTCAGACAAATGCTACTAT | 2068 | | |
| Qy | 661 | TTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAGTCTTCA | 720 | | |
| Db | 2069 | TTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAGTCTTCA | 2128 | | |
| Qy | 721 | CATCTTGTTTTATAAACAATAGAGGAACAGCAATGGATAAAAAACAGATGGTAGG | 780 | | |
| Db | 2129 | CATCTTGTTTTATAAACAATAGAGGAACAGCAATGGATAAAAAACAGATGGTAGG | 2188 | | |
| Qy | 781 | AGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGAATGGGCTG | 840 | | |
| Db | 2189 | AGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAATGAATGAATGGGCTG | 2248 | | |
| Qy | 841 | GATGGACATCTCCAGACTCAAAAATTTGAAAGCTGACAGCCGCGATAACTGGGGTCA | 900 | | |
| Db | 2249 | GATGGACATCTCCAGACTCAAAAATTTGAAAGCTGACAGCCGCGATAACTGGGGTCA | 2308 | | |
| Qy | 901 | GGCCATGGCCAGGAGAGACTGTGCTGGTGTGATTTATGCTGGCGAGTGGAAAGATTTC | 960 | | |
| Db | 2309 | GGCCATGGCCAGGAGAGACTGTGCTGGTGTGATTTATGCTGGCGAGTGGAAAGATTTC | 2368 | | |
| Qy | 961 | CAATGTGAGACGCTCAATACTTATTTGCGAAAAAGACAGAGGAGACAGTACTGTCACT | 1020 | | |
| Db | 2369 | CAATGTGAGACGCTCAATACTTATTTGCGAAAAAGACAGAGGAGACAGTACTGTCACT | 2428 | | |

| | | | | | |
|----|------|---|------|--|--|
| Qy | 1021 | GCATTATAACGGAAGTGGATGCGATGCGATGAGCAAAATTTTCAGCTCTCAAGGCAAGG | 1080 | | |
| Db | 2429 | GCATTATAACGGAAGTGGATGCGATGAGCAAAATTTTCAGCTCTCAAGGCAAGG | 2488 | | |
| Qy | 1081 | ACACTCCTTTCTAATGTCATCACCTTCTCATCAGATTGAAAAAAGAAAGACTGAAAA | 1140 | | |
| Db | 2489 | ACACTCCTTTCTAATGTCATCACCTTCTCATCAGATTGAAAAAAGAAAGACTGAAAA | 2547 | | |
| Qy | 1141 | CCAATTACTGAAAAAATTCACAGCTAGTGTCTTTTACCATCGTCATTACCCAAAGAC | 1200 | | |
| Db | 2548 | CCAATTACTGAAAAAATTCACAGCTAGTGTCTTTTACCATCGTCATTACCCAAAGAC | 2607 | | |
| Qy | 1201 | TTGGGAACATAAATGTTCCCGCAGGCTGATATGCTGATTTTCATTTGTCACACTGGA | 1260 | | |
| Db | 2608 | TTGGGAACATAAATGTTCCCGCAGGCTGATATGCTGATTTTCATTTGTCACACTGGA | 2667 | | |
| Qy | 1361 | TCACATAGATTCTCCTCGTCCAGTAACCGTCCGATATACAAATATATCTTCCAAAGTA | 1320 | | |
| Db | 2668 | TCACATAGATTCTCCTCGTCCAGTAACCGTCCGATATACAAATATATCTTCCAAAGTA | 2727 | | |
| Qy | 1371 | TGGAACACTCCAATCAGAAAAAGGTTATCAT | 1351 | | |
| Db | 2728 | TGGAACACTCCAATCAGAAAAAGGTTATCAT | 2758 | | |

RESULT 7

AB078434

LOCUS

DEFINITION

AB078434

AB078434

AB078434.1

GI:21901968

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

AB078434

Mus musculus CL-P1 mRNA for collectin placenta 1, complete cds.

2637 bp

mRNA

linear

ROD 18-JUL-2002

Mus musculus

Mus musculus

Mus musculus tissue_lib:Liver cdna to mRNA.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1

Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H., Sakai,Y., Fukuoh,A., Sakamoto,T., Itabe,H., Suzutani,T., Ogasawara,M., Yoshida,I. and Wakamiya,N.

cDNA cloning of mouse CL-P1 gene

Unpublished

2 (bases 1 to 2637)

Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H., Sakai,Y., Fukuoh,A., Sakamoto,T. and Wakamiya,N.

Direct Submission

Submitted (18-JAN-2002) Katsuki Ohtani, Asahikawa Medical College, Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa, Hokkaido 078-8510, Japan (E-mail: ohtani@asahikawa-med.ac.jp, Tel:81-166-68-2393, Fax:81-166-68-2399)

Location/Qualifiers

1..2637

/organism="Mus musculus"

/db_xref="taxon:10090"

/tissue_lib="Liver"

1..2637

/gene="CL-P1"

92..2320

/gene="CL-P1"

/codon_start=1

/product="collectin placenta 1"

/protein_id="BAC05523.1"

/db_xref="GI:21901968"

/translation="MKDFAEEVQSFYKRFQIQGTQCTKCKNNWALKFSLVLY ILCALLTITVAILGKVKVERKNDVDTGMETSHQYTDNKLTAVESDLKKLGGQAKKAL STNSLSTFRSDILDLRQQLQEIETKSKNKDTLEKLANGDSLVDRLQSQKLTQNN SFLITVTKNTLQAYNGVTVLQDDTSVQNLQSQVIMNLNLNTQVQRN LLSNLQOSVDDTSIAIORIKNDLQVLOQFLOAKKDDDLKKEVQSLOTIAANNSALA KANNITLEDMSQLSSFTGQMDNTTISQANEQSLKDLQDLKDKTENTRTAVKFSOLEE RFQVFTDIWNISISYTHAHLRTLSNLDVTRCTCTDLTRHTDOLTSINLNLVNI RLDSISLRMQDDMMRSLKDTVEANLVSVMMEKMLVDSKHGQLTKNFTILQGPGRGP KDRGSGQPPGPTGNKGQKGEKGPFGPAGERGTIGVPVGPFGERSGSGKSGQSPK

SFLITVYVNTLQAYNGYVYNLQDDTNVLQGNLQSQMYNSQSVVYNNLNLNLTQVQORN
LISNLQSDVDTLSIAORIKNDFONLQOVFLQAKKDDWLKKEKYOSVSTQTLAANSALA
KANNDNTSISQSESTQGMNDNTTISQANEQSLKDLQDLKHDTENRTAVKFSQLEE
RFQVETDIVIINSYTAHLRLTLNLNDVMTCTDTRLTRTDLLTSLNNLTVNI
RLDSISLRMOQDMRSKLDTEVANLSVVMEMKLVDSKHGOLIKNFTLLOQPPGRGP
KGRDSQSPGPTNKQKQKBEKBPBPAGERTGPGVPPGERSKSGSKSQGPK
GSRSPKPGPGSGDPPGPKDGLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
LPVPGMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
KEILEDKLFCESSHLVFINREEQWIKKHTVGRSHWIGLTDSPQSEWKWLDG
SPVDYKNKAGQPDNWSGHPGDEDCAGLIYAGQWINDQFCEINNFICEKREAVPSS
IL"

| | | | | | |
|----------------------------|------|--|-------|-------|-------|
| polyA_signal | | 3246..3251 | | | |
| BASE COUNT | | 1013 a | 757 c | 759 g | 762 t |
| ORIGIN | | | | | |
| Query Match | | 65.5%; Score 887.4; DB 10; Length 3291; | | | |
| Best Local Similarity | | 81.5%; Pred. No. 1.4e-195; | | | |
| Matches 1105; Conservative | | 0; Mismatches 231; Indels 19; Gaps 6; | | | |
| QY | 1 | ATGCAACAAGATTGATGAGTTCGAGTTAGACACTCAAGTAGCCAACTTATCAGTGATT | 60 | | |
| DB | 1277 | ATGCACGAACATGATGAGGTCAAAGTTAGACACTGAAGTGCCCACTTATCAGTGGTT | 1336 | | |
| QY | 61 | ATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTTCAGCTCATCAAGAATTTTACAATA | 120 | | |
| DB | 1337 | ATGGAAGAGATGAACATGGTTGACTCCAAGCACGGTCAGTCTCAGAGACTTTACCATT | 1396 | | |
| QY | 121 | CTACAAGGTCCACGGGCCCGCCAGGGTCCAAAGAGTGACAGAGATCCCAGGGACCCCT | 180 | | |
| DB | 1397 | CTACAAGGTCTCTTGGSCCCAGAGGTCCAAAAGGTGCACAGAGGATCTCAGGGACCACT | 1456 | | |
| QY | 181 | GGCCCAACTGGCAACAAGGACAGAAAGAGAGAGGGGAGCCCTGGACCACTTGGGCCCT | 240 | | |
| DB | 1457 | GGTCCAACTGGCAACAAGGACAGAAAGAGAGAGGGGAGCCCTGGTCCACTTGGCCCT | 1516 | | |
| QY | 241 | GCSSGTTGAGAGAGCCCAATTGGACAGCTGTGTCGCCCGGAGAGCGTGGCGGCAAGGA | 300 | | |
| DB | 1517 | GCSSGTTGAGAGGGGACAAATTGGACAGTTCGGCCCTCTTGAGAGCGTGGCAGCAAGGA | 1576 | | |
| QY | 301 | TCTAAAGGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCTCTGGGAAAGCCGGGCCCTCAG | 360 | | |
| DB | 1577 | TCCAAGGCTCACAGGTCCTCAAGAGTCTCTGGGTTCCTCCAGGGAGAGCTTGGCCCTCAA | 1636 | | |
| QY | 361 | GGCCCACTGGGACCCAGGCCCGCCGGGCCACCAAGGCAAGAGGAGTCCCGGCCCT | 420 | | |
| DB | 1637 | GGACCTAGTGGGACCCAGGACCAAGGTCCACAGGCAAGGATGGACTCCCTTGGCCCT | 1696 | | |
| QY | 421 | CAGGCGCTCTCTGGCTTCCAGGACTTCCAGGACCGTTGGGGAGCGCTGGGTGCTTGA | 480 | | |
| DB | 1697 | CAGGCGCTCTCTGGCTTCCAGGACTTACAGGCACTGTGGTGGAGTGGAGTACCTTGA | 1756 | | |
| QY | 481 | CTTCGGGAGCTGCAGGCTTTCCTGGGTACAGGATGCCAGGCGCCCAAGGGCCCCCCC | 540 | | |
| DB | 1757 | CTTCGGGAGTTCAGAGCTTGCAGGGTGCAGGATGCTTGGCCCTAAGGAGCACTT | 1816 | | |
| QY | 541 | GGCCCTCTTGGCCATCAGGAGCGGTGGTGCCTTGGCCCTCAGCAATGAGCCAAACCCG | 600 | | |
| DB | 1817 | GGCCCTCCAGGCGCTCAGGACCAATGGAGCCATTTGGCTCTGCAGATGAACCAACCCCA | 1876 | | |
| QY | 601 | GCACGGGAGGCAATGGCTCCCGCTCTCAGTGAAGAACTTTCACAGACAAATGCTACTAT | 660 | | |
| DB | 1877 | GCATCAGAGGTCAACGGATGTCGCGCTCTCAGTGAAGAACTTTCACAGATAAATGCTACTAT | 1936 | | |
| QY | 661 | TTTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAAGCAAGTCTTCA | 720 | | |
| DB | 1937 | TTTTTCATTTGAAAAAGAAATTTTGAAGATGCTAAGCTTTTCTGTGAAGCAAACTCTTCC | 1996 | | |
| QY | 721 | CATCTGTTTTTCATAAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGG | 780 | | |
| DB | 1997 | CATCTGTTTTTCATAAACTCAAGAGAAACAGCAATGGATAAAAAACAGATACCGTGGGG | 2056 | | |
| QY | 781 | AGAGAGAGCCACTGGATCGGCCCTCACAGACTCAGAGCGTGAATAATGAATGGAAGTGGCTG | 840 | | |

| | | | | | |
|----|------|---|------|--|--|
| Db | 2057 | AGAGAAAGCCATTGGATCGGCTTCAGACTCAGAACAGAGAAAGCAATGGAAGTGGCTA | 2116 | | |
| Qy | 841 | GATGGACATCTCCAGACTACAAAAATTCGAAAGCTGGACACCGGATACACTGGGCTCAT | 900 | | |
| Db | 2117 | GACGGTCACTGTTGATTACAAAACTGGAAGCTGGACAAACAGGATACACTGGGGCACT | 2176 | | |
| Qy | 901 | GGCCATGGCCAGGAGAGACTGTGCTGGTGTGATTATGCTGGGAGTGGACGATTTTC | 960 | | |
| Db | 2177 | GGCCATGGCCAGGAGAGACTGTGCTGGTGTGATTATGCTGGGAGTGGACGATTTTC | 2236 | | |
| Qy | 961 | CAATGTGAAGACCTCAATTAACCTTTCGCGAAAGAGACAGGAGACAGTACTGTCATCT | 1020 | | |
| Db | 2237 | CAGTGTGATGAATCAATAAATTCATTGTGAGAAAGGAAAGGAGGACATCATCATCC | 2296 | | |
| Qy | 1021 | GCATTATAACGACCTGTGATGGATCAGATGAGCAAAATTTTCA--GCTCTCAAAAGGCAAA | 1078 | | |
| Db | 2297 | ATATTATACACATGATATATAGCAGAAACATATTTTCTGATGCTCTGAAAGCCGAA | 2356 | | |
| Qy | 1079 | GGACACTCTTTTCTTAATTCACCTTCTCATCAGATTGAAAAAAGGAGGAGGAGGAGG | 1138 | | |
| Db | 2357 | GAATGCTGTTTTCATTCATCA--CTTCTCACCAGATTG--AATGGAAGGAGGCTCGAA | 2413 | | |
| Qy | 1139 | AACCAATTAAGTGAAGAAATTCAGACGCTAGTGTGTTTACCACGCTGATACCCAAAG | 1198 | | |
| Db | 2414 | AAGTAGTTATTCAAAATAAATGGACACCTACTG-----CACAATAACCCCAAGG | 2461 | | |
| Qy | 1199 | ACTTGGGAACTAAATG--TTCCCGCAGGGTATGCTGATTT--TCATTCGACATGGAC | 1256 | | |
| Db | 2462 | ACTAGGGGCTAAATGCTCCCGCAGTTGATATATGTTTATTTCCCAAGTGTACAAATGGAC | 2521 | | |
| Qy | 1257 | TGAATCAGATAGATTCTCTCGTCACTAACCGTGCATTATATACAAATTTATGCTTTCCAA | 1316 | | |
| Db | 2522 | TGAATCGCATAGATTCTCTCAGCCATTACCATAGATTATGCAAGATATATCTTTCCA | 2581 | | |
| Qy | 1317 | AGTATGGAACACTTCAATCAGAAAAAGGTTATCAT | 1351 | | |
| Db | 2582 | AATATGGAATGCTCCAATCAGAAAAAGACTATCAT | 2616 | | |

RESULT 9
AB052103 Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin type II, complete cds.
ACCESSION AB052103
VERSION AB052103.1 GI:13365552
KEYWORDS Homo sapiens cdna to mRNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (sites)
AUTHORS Nakamura, K., Funakoshi, H., Miyamoto, K., Tokunaga, F. and Nakamura, T.
TITLE Molecular cloning and functional characterization of a human scavenger receptor with C-type lectin (SRCL), a novel member of a scavenger receptor family
JOURNAL Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
MEDLINE 21092718
REFERENCE 2 (bases 1 to 4330)
AUTHORS Nakamura, K. and Nakamura, T.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-2000) Kenji Nakamura, Osaka University Graduate School of Medicine, Division of Biochemistry, Biomedical Research Center; 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail: knakamur@onbich.med.osaka-u.ac.jp, Tel: 81-6-6879-3783, Fax: 81-6-6879-3789)
FEATURES
source 1. 4330
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene 1. 4330
/gene="SRCL"

CDS

3...1871
/gene="SRCL"
/codon_start=1
/product="scavenger receptor with C-type lectin type II"
/protein_id="BAB39148.1"
/db_xref="GI:1336553"
/translation="MKDDFAEEVEVPFGRKRGIEHGTQCPKCHNNWALKFSILLY
ILCALLTTTVAIIYKVVRKMDNTGGMETSRQTVDDKLTAVESDLKLGQDTGKKAI
STNSLSFRDILDLROOLREITKTSKNDKTLKLOASGDALVDROSOLKTKLENN
SFLITVNTKTLQAVNGVYVTLQODTSVLQGNLQNMVSHNVVIMNLLNLTQVQORN
LITNLSRVDDTSQAIQRIKDNFQNLQVFLQAKKDDWLKQVSLQTLAANSALA
KANNDLSDMSQNSLSTQGMENITTSQANEQNLKLDLHKDAENRTAKFNOLKEE
RFQFETDVIINISYTAHLLRTLSNNEVRTCTDTLTKHTDDTLSSNNTLANI
RLDSVLSRMQDLARSLRDTVEANLSVMEEMKLVDSKGLIKNFITLQGPGRGP
RGDSGSGPPTNGKQGEKGPBPAGRGPIGPAGPGRGKSGKSGQGP
GSRSPGKPGQSGDGPBPKEGLPCQGPQPGQGLQGTGVEGVPBPGLPG
LPGVGPBPGRPGPPGPPGSAVPLALNPTPAPEDNSKSPSLQPGGGSACA"
BASE COUNT 1186 a 1065 c 1035 g 1044 t
ORIGIN

| | | | | | | | |
|-----------------------|------|--|---------------------|---------------|--------------|--|--|
| Query Match | | 53.3% | Score 722.4; | DB 9; | Length 4330; | | |
| Best Local Similarity | | 99.7% | Pred. No. 3.2e-157; | | | | |
| Matches 734; | | Conservative | 0; | Mismatches 1; | Indels 1; | | |
| QY | 616 | GGCTGCCCGCCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA | 675 | | | | |
| Db | 3414 | GGCTGCCCGCCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA | 3473 | | | | |
| QY | 676 | GAATTTTGGAGTGCRAAGCTTTCTGTGAAGACAACTCTTCACATCTTCTTTTCATA | 735 | | | | |
| Db | 3474 | GAATTTTGGAGTGCRAAGCTTTCTGTGAAGACAACTCTTCACATCTTCTTTTCATA | 3533 | | | | |
| QY | 736 | AACACTAGAGAGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGG | 795 | | | | |
| Db | 3534 | AACACTAGAGAGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGG | 3593 | | | | |
| QY | 796 | ATCGGCCTCACAGACTCAGAGCGTGAATGAATGAAGTGGCTGGATGGGACATCTCCA | 855 | | | | |
| Db | 3594 | ATCGGCCTCACAGACTCAGAGCGTGAATGAATGAAGTGGCTGGATGGGACATCTCCA | 3653 | | | | |
| QY | 856 | GACTACAAAAATTGGAAAGCTGGACAGCGGATAACTTGGGGTCATGGCCATGGGCCAGGA | 915 | | | | |
| Db | 3654 | GACTACAAAAATTGGAAAGCTGGACAGCGGATAACTTGGGGTCATGGCCATGGGCCAGGA | 3713 | | | | |
| QY | 916 | GAAGACTGTGCTGGTGTGATTATGCTGGGCAGTGGAGCAATTTCCAAATGTGAAGACGTC | 975 | | | | |
| Db | 3714 | GAAGACTGTGCTGGTGTGATTATGCTGGGCAGTGGAGCAATTTCCAAATGTGAAGACGTC | 3773 | | | | |
| QY | 976 | AATAACTTTCATTTTCGAAAAACAGACGGGAGACAGTACTGTCTATCTGCATTTATAACGGACT | 1035 | | | | |
| Db | 3774 | AATAACTTTCATTTTCGAAAAACAGACGGGAGACAGTACTGTCTATCTGCATTTATAACGGACT | 3833 | | | | |
| QY | 1036 | GTGATGGGATACATGAGCAAAATTTTCAGTCTCTCAAAGCAAGGACACTCCCTTCTTAAT | 1095 | | | | |
| Db | 3834 | GTGATGGGATACATGAGCAAAATTTTCAGTCTCTCAAAGCAAGGACACTCCCTTCTTAAT | 3893 | | | | |
| QY | 1096 | TGCATCACCTTCTCATCAGATTGAAAAAAGCACTGAAAAACCAATTAATCTGAAAAA | 1155 | | | | |
| Db | 3894 | TGCATCACCTTCTCATCAGATTGAAAAAAGCACTGAAAAACCAATTAATCTGAAAAA | 3952 | | | | |
| QY | 1156 | AAATTGACAGCTAGTGTGTTTTTACCATCCGTCATTAACCCAAAGACTTGGGAACATAAATG | 1215 | | | | |
| Db | 3953 | AAATTGACAGCTAGTGTGTTTTTACCATCCGTCATTAACCCAAAGACTTGGGAACATAAATG | 4012 | | | | |
| QY | 1216 | TTCCCCAGGCTGATATGCTGATTTTCATTTGTCACATGGACTGAATCAGATATCTCC | 1275 | | | | |
| Db | 4013 | TTCCCCAGGCTGATATGCTGATTTTCATTTGTCACATGGACTGAATCAGATATCTCC | 4072 | | | | |
| QY | 1276 | TCCGTCAAGTAACCGTGGCAATATACAAATATGTCTTCCAAAGTATGGACACTCCAATC | 1335 | | | | |
| Db | 4073 | TCCGTCAAGTAACCGTGGCAATATACAAATATGTCTTCCAAAGTATGGACACTCCAATC | 4132 | | | | |
| QY | 1336 | AGAAAAAGGTTATCAT | 1351 | | | | |

Db 4133 AAAAAAGGTTATCAT 4148

RESULT 10
AC016128/c
LOCUS

DEFINITION Homo sapiens chromosome 18 clone RP11-324G2 map 18, WORKING DRAFT
SEQUENCE, 8 unordered pieces.

ACCESSION AC016128
VERSION AC016128.4 GI:10046526

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 169088)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

JOURNAL Title
REFERENCE 2 (bases 1 to 169088)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castelle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Leloczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McSwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meidrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye W.J., Zimmer,A. and Zody,M.

TITLE Direct Submission
JOURNAL Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome

COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 9, 2000 this sequence version replaced gi:5649269.

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L999

Center clone name: 324_G2
----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731
Consensus quality: 16522 bases at least Q40

Consensus quality: 167248 bases at least Q30
Consensus quality: 167996 bases at least Q20

Insert size: 177000; agarose-fp
Insert size: 168388; sum-of-contigs

Quality coverage: 6.2 in Q20 bases; agarose-fp
Quality coverage: 6.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.

* 1 17943: contig of 17943 bp in length
* 17944 18043: gap of 100 bp
* 18044 22424: contig of 4381 bp in length

* 22425 22524: gap of 100 bp
* 22525 38094: contig of 15570 bp in length
* 38095 38194: gap of 100 bp
* 38195 49220: contig of 11026 bp in length
* 49221 49320: gap of 100 bp
* 49321 66394: contig of 17074 bp in length
* 66395 66494: gap of 100 bp
* 66495 91692: contig of 25198 bp in length
* 91693 91792: gap of 100 bp
* 91793 120869: contig of 29077 bp in length
* 120870 120869: gap of 100 bp
* 120970 169088: contig of 48119 bp in length.

FEATURES

source
1. .169088
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18"
/chromosome="18"
/clone="RP11-32492"
/clone_lib="RPC1-11 Human Male BAC"
1. .17943
/note="assembly_fragment
clone_end:SP6
vector_side:left"
18044. .22424
/note="assembly_fragment"
22525. .38094
/note="assembly_fragment"
38195. .49220
/note="assembly_fragment"
49321. .66394
/note="assembly_fragment"
66495. .91692
/note="assembly_fragment"
91793. .120869
/note="assembly_fragment"
120970. .169088
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 48103 a 33537 c 34088 g 52660 t 700 others
ORIGIN

Query Match 36.1%; Score 488.8; DB 2; Length 169088;
Best Local Similarity 99.6%; Pred. No. 9.2e-103;
Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 126 AGGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCAGGACCCCTGGGCC 185
Db 153817 AGGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCAGGACCCCTGGGCC 153758
QY 186 AACTGGCAACAAAGGACAGAGAGAGAGAGGGGAGCTGGACCACTGGCCCTGGGG 245
Db 153757 AACTGGCAACAAAGGACAGAGAGAGAGAGGGGAGCTGGACCACTGGCCCTGGGG 153698
QY 246 TGAGAGAGGCCCAATTGGACCACTGGTTCCTCCCGGAGAGCTGGCGCAAAAGGATCTAA 305
Db 153697 TGAGAGAGGCCCAATTGGACCACTGGTTCCTCCCGGAGAGCTGGCGCAAAAGGATCTAA 153638
QY 306 AGGCTCCAGGGCCCGCAAAAGCTCCGTTGTTTCCCTGGGAAGCCCGGCCCTCAGGGCCC 365
Db 153637 AGGCTCCAGGGCCCGCAAAAGCTCCGTTGTTTCCCTGGGAAGCCCGGCCCTCAGGGCTC 153578
QY 366 CAGTGGGACCCAGCGCCCGCCGACCAAGGAGGAGCTCCCGGCCCTCAGGG 425
Db 153577 CAGTGGGACCCAGCGCCCGCCGACCAAGGAGGAGCTCCCGGCCCTCAGGG 153518
QY 426 CCCTCCTGGCTTCAGGAGCTTCAGGACACCGTTGGGAGAGCTGGGGTGCCTGGACCTCG 485
Db 153517 CCCTCCTGGCTTCAGGAGCTTCAGGACACCGTTGGGAGAGCTGGGGTGCCTGGACCTCG 153458
QY 486 GGGACTGCCAGGCTTGCTGGGGTACAGGCATGCCAGGCCCAAGGGCCCGCCCGGCC 545
|||||

Db 153457 GGGACTGCCAGGCTTGCTGGGGTACCAGGCATGCCAGGCCCGCCCGGCCCGGCC 153398
QY 546 TCCTGGCCCATCAGAGCGGTGGTCCCTGGCCCTGCAGATGAGCAACCCCGGCACC 605
|||||
Db 153397 TCCTGGCCCATCAGAGCGGTGGTCCCTGGCCCTGCAGATGAGCAACCCCGGCACC 153338
QY 606 GGAGGACAATGG 617
|||||
Db 153337 GGAGGACAATGG 153326
|||||
RESULT 11
AP000915/c 188439 bp DNA linear PRI 27-APR-2002
LOCUS Homo sapiens genomic DNA, chromosome 18p clone:RP11-720L2, complete
DEFINITION sequences.
ACCESSION AP000915
VERSION AP000915.5 GI:20334314
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-720L2.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (1999)
REFERENCE 2 (bases 1 to 188439)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Apr 26, 2002 this sequence version replaced gi:9188470.
FEATURES
Location/Qualifiers
source
1..188439
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p"
/clone="RP11-720L2"
BASE COUNT 54003 a 41134 c 40027 g 53275 t
ORIGIN
Query Match 36.1%; Score 488.8; DB 9; Length 188439;
Best Local Similarity 99.6%; Pred. No. 9.3e-103;
Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 126 AGGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCAGGACCCCTGGGCC 185
Db 44938 AGGTCCACCGGGCCCGAGGGTCCAAAGAGGTGACAGAGATCCAGGACCCCTGGGCC 44879
QY 186 AACTGGCAACAAAGGACAGAGAGAGAGAGGGGAGCTGGACCACTGGCCCTGGCGG 245
Db 44878 AACTGGCAACAAAGGACAGAGAGAGAGAGGGGAGCTGGACCACTGGCCCTGGCGG 44819
QY 246 TGAGAGAGGCCCAATTGGACCACTGGTTCCTCCCGGAGAGCGTGGCGCAAAAGGATCTAA 305
Db 44818 TGAGAGAGGCCCAATTGGACCACTGGTTCCTCCCGGAGAGCGTGGCGCAAAAGGATCTAA 44759
QY 306 AGGCTCCAGGGCCCGCAAAAGCTCCCGTTGTTTCCCTGGGAAGCCCGGCCCTCAGGGCCC 365
Db 44758 AGGCTCCAGGGCCCGCAAAAGCTCCCGTTGTTTCCCTGGGAAGCCCGGCCCTCAGGGCTC 44699
QY 366 CAGTGGGACCCAGCGCCCGCCGACCAAGGAGGAGCTCCCGGCCCTCAGGG 425
Db 44698 CAGTGGGACCCAGCGCCCGCCGACCAAGGAGGAGCTCCCGGCCCTCAGGG 44639
QY 426 CCCTCCTGGCTTCAGGAGCTTCAGGACACCGTTGGGAGAGCTGGGGTGCCTGGACCTCG 485
|||||


```
|||||
Db 44638 CCTCTGCTTCCAGGACTTCAGGGACACCGTTGGGAGCGTGGGGTGCCTTGGACCTCG 44579
QY 486 GGGACTGCCAGGCTTGCCTGGGGTACCAAGGATGCGAGGCCCAAGGCCGCCCGCGGCC 545
|||||
Db 44578 GGGACTGCCAGGCTTGCCTGGGGTACCAAGGATGCGAGGCCCAAGGCCGCCCGCGGCC 44519
QY 546 TCCTGGCCCATCAGAGCGGTGGTGCCTGGCCCTGCGAGATGAGCAACCCCGGCACC 605
|||||
Db 44518 TCCTGGCCCATCAGAGCGGTGGTGCCTGGCCCTGCGAGATGAGCAACCCCGGCACC 44459
QY 606 GGAGGACATGG 617
Db 44458 GGAGGACATGG 44447

RESULT 12
LOCUS AC024368 71044 bp DNA linear HTG 28-FEB-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-179K3 map 11, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC024368
VERSION AC024368.1 GI:7108157
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 71044)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RP11-179K3
2 (bases 1 to 71044)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavskiy,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivar,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6145
Center clone name: 179_K_3
-----
* NOTE: This record contains 90 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
```

```
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 698: contig of 698 bp in length
* 699 798: gap of 100 bp
* 799 1495: contig of 697 bp in length
* 1496 1595: gap of 100 bp
* 1596 2284: contig of 689 bp in length
* 2285 2384: gap of 100 bp
* 2385 3056: contig of 672 bp in length
* 3057 3156: gap of 100 bp
* 3157 3851: contig of 695 bp in length
* 3852 3951: gap of 100 bp
* 3952 4627: contig of 676 bp in length
* 4628 4727: gap of 100 bp
* 4728 5409: contig of 682 bp in length
* 5410 5509: gap of 100 bp
* 5510 6183: contig of 674 bp in length
* 6184 6283: gap of 100 bp
* 6284 6984: contig of 701 bp in length
* 6985 7084: gap of 100 bp
* 7085 7768: contig of 684 bp in length
* 7769 7868: gap of 100 bp
* 7869 8550: contig of 682 bp in length
* 8551 8650: gap of 100 bp
* 8651 9352: contig of 702 bp in length
* 9353 9452: gap of 100 bp
* 9453 10140: contig of 688 bp in length
* 10141 10240: gap of 100 bp
* 10241 10922: contig of 682 bp in length
* 10923 11022: gap of 100 bp
* 11023 11709: contig of 687 bp in length
* 11710 11809: gap of 100 bp
* 11810 12511: contig of 702 bp in length
* 12512 12611: gap of 100 bp
* 12612 13304: contig of 693 bp in length
* 13305 13404: gap of 100 bp
* 13405 14095: contig of 691 bp in length
* 14096 14195: gap of 100 bp
* 14196 14881: contig of 686 bp in length
* 14882 14981: gap of 100 bp
* 14982 15674: contig of 693 bp in length
* 15675 15774: gap of 100 bp
* 15775 16460: contig of 686 bp in length
* 16461 16560: gap of 100 bp
* 16561 17250: contig of 690 bp in length
* 17251 17350: gap of 100 bp
* 17351 18040: contig of 690 bp in length
* 18041 18140: gap of 100 bp
* 18141 18836: contig of 696 bp in length
* 18837 18936: gap of 100 bp
* 18937 19621: contig of 685 bp in length
* 19622 19721: gap of 100 bp
* 19722 20425: contig of 704 bp in length
* 20426 20525: gap of 100 bp
* 20526 21224: contig of 699 bp in length
* 21225 21324: gap of 100 bp
* 21325 22017: contig of 693 bp in length
* 22018 22117: gap of 100 bp
* 22118 22815: contig of 698 bp in length
* 22816 22915: gap of 100 bp
* 22916 23609: contig of 694 bp in length
* 23610 23709: gap of 100 bp
* 23710 24402: contig of 693 bp in length
* 24403 24502: gap of 100 bp
* 24503 25191: contig of 689 bp in length
* 25192 25291: gap of 100 bp
* 25292 25981: contig of 690 bp in length
* 25982 26081: gap of 100 bp
```


| | | | | |
|-----------------------|-----------------|-------------------|-----------|---------------|
| Query Match | 35.8% | Score 484.6; | DB 2; | Length 71044; |
| Best Local Similarity | 98.2%; | Pred. No. 8e-102; | | |
| Matches 490; | Conservative 0; | Mismatches 9; | Indels 0; | Gaps 0; |

| | | | |
|----|-------|---|-------|
| Qy | 119 | TACTACAAGGTCCACCGGSCCCAGGGCTCCAAGAGGTGACAGAGGATCCACGAGCACCC | 178 |
| Db | 43599 | TAGCAAAAGTTCACCGGSCCCAGGGTCCAAGAGGTGACAGAGATCCACGAGCACCC | 43658 |
| Qy | 179 | CTGGCCCAACTGGCAACAAGGACAGAAGAGAGAGGGGAGGCTTGACCAACCTGGCC | 238 |
| Db | 43659 | CTGGCCCAACTGGCAACAAGGACAGAAGAGAGAGAGGGGAGGCTTGACCAACCTGGCC | 43718 |
| Qy | 239 | CTGGGGGTGACAGAGGCCCAATTGGACAGAGTGTGCCCCCGGAGAGCGTGGCGCAAG | 298 |
| Db | 43719 | CTGGGGGTGACAGAGGCCCAAGGAGGAGGCTGTGCCCCCGGAGAGCGTGGCGCAAG | 43778 |
| Qy | 299 | GATCTAAAGGCTCCACGGGCCCAAGAGCTCCCGTGTTCCTCTGGGAAGCCCGGCCCTC | 358 |
| Db | 43779 | GATCTAAAGGCTCCACGGGCCCAAGAGCTCCCGTGTTCCTCTGGGAAGCCCGGCCCTC | 43838 |
| Qy | 359 | AGGGCCCACTGGGAGACCCAGGCCCCCGGGCCCAACGAGCAAAAGAGGACTCCCCGACC | 418 |
| Db | 43839 | AGGGCTCCAGTGGGAGACCCAGGCCCCCGGGCCCAACGAGCAAAAGAGGACTCCCCGACC | 43898 |
| Qy | 419 | CTCAGGGCCCTCTGGGCTTCCAGGAGCTTCAGGGACACCTGTTGGGAGAGCCTGGGGTGCCTG | 478 |
| Db | 43899 | CTCAGGGCCCTCTGGGCTTCCAGGAGCTTCAGGGACACCTGTTGGGAGAGCCTGGGGTGCCTG | 43958 |
| Qy | 479 | GACCTCGGGGACTGCCAGGCTTGCTGGGTACCAAGGATGCCAGGCCCAAGGGSCCCC | 538 |
| Db | 43959 | GACCTCGGGGACTGCCAGGCTTGCTGGGTACCAAGGATGCCAGGCCCAAGGGSCCCC | 44018 |
| Qy | 539 | CCGGCCCTCTGGCCCATCAGAGCGGTGTGCCCTTGCCCTCTGCAGAAATGAGCCAAGCC | 598 |
| Db | 44019 | CCGGCCCTCTGGCCCATCAGAGCGGTGTGCCCTTGCCCTCTGCAGAAATGAGCCAAGCC | 44078 |
| Qy | 599 | CGGCACCGGAGACAATGG | 617 |
| Db | 44079 | CAGACCGGAGACAATGG | 44097 |

```

RESULT 13
AP001022      187635 bp      DNA      linear      HTG 30-MAY-2000
LOCUS      Homo sapiens chromosome 18 clone RP11-815L4 map 18p11.3, WORKING
DEFINITION      DRAFT SEQUENCE, 33 unordered pieces.
ACCESSION      AP001022
VERSION      AP001022.2 GI:8117692
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-815L4.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 187635)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 187,635 genomic DNA of 18p11.3
Published Only in DataBase (2000)
2 (bases 1 to 187635)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (05-JAN-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagami-hara, Kanagawa 228-8555,
Japan (E-mail: hattori@gscc.riken.go.jp,
URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923,
Fax: 81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6997772.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web page: http://hgp.gsc.riken.go.jp/

COMMENT

```

Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-815L4
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167192 bases at least Q40
Consensus quality: 177020 bases at least Q30
Consensus quality: 182105 bases at least Q20
Insert size: 18435; sum-of-contigs
Quality coverage: 4.35x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved
1 21332 contig of 21332 bp in length
21433 42072 contig of 20640 bp in length
42173 54336 contig of 12164 bp in length
54437 66766 contig of 12330 bp in length
66867 77849 contig of 10983 bp in length
77950 87798 contig of 9849 bp in length
87899 99510 contig of 11612 bp in length
99611 107273 contig of 100 bp
107374 13189 contig of 7663 bp in length
13290 120559 contig of 5091 bp in length
120660 126468 contig of 3015 bp in length
126569 131619 contig of 4207 bp in length
131720 136810 contig of 5091 bp in length
136911 139925 contig of 3015 bp in length
140026 144232 contig of 4207 bp in length
14333 148902 contig of 4570 bp in length
148903 152918 contig of 3316 bp in length
152919 153018 contig of 100 bp
153019 156399 contig of 3381 bp in length
156400 156499 contig of 100 bp
156500 158933 contig of 2434 bp in length
158934 159033 contig of 100 bp
159034 161884 contig of 2851 bp in length
161885 164301 contig of 2317 bp in length
164302 164401 contig of 100 bp
164402 166300 contig of 1899 bp in length
166301 166400 contig of 100 bp
166401 168580 contig of 2180 bp in length
168581 168680 contig of 100 bp
168681 171270 contig of 2590 bp in length
171271 171370 contig of 100 bp
171371 174092 contig of 2722 bp in length
174093 174192 contig of 100 bp
174193 176095 contig of 1903 bp in length
176096 176195 contig of 100 bp
176196 177890 contig of 1695 bp in length
177891 177990 contig of 100 bp
177991 179786 contig of 1796 bp in length
179787 179886 contig of 100 bp
179887 181478 contig of 1592 bp in length
181479 181578 contig of 100 bp
181579 182895 contig of 1317 bp in length
182896 182996 contig of 100 bp
182996 184841 contig of 1846 bp in length
184842 184941 contig of 100 bp
184942 186345 contig of 1404 bp in length
186346 186445 contig of 100 bp
186446 187635 contig of 1190 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1 21332: contig of 21332 bp in length
21433 21432: gap of 100 bp
21433 42072: contig of 20640 bp in length
42073 42172: gap of 100 bp
42073 54336: contig of 12164 bp in length
54337 54436: gap of 100 bp
54437 66766: contig of 12330 bp in length
66767 66866: gap of 100 bp
66867 77849: contig of 10983 bp in length
77850 77949: gap of 100 bp

* 77950 87798: contig of 9849 bp in length
* 87799 87898: gap of 100 bp
* 87899 99510: contig of 11612 bp in length
* 99511 99610: gap of 100 bp
* 99611 107273: contig of 7663 bp in length
* 107274 107373: gap of 100 bp
* 107374 113189: contig of 5816 bp in length
* 113190 113289: gap of 100 bp
* 113290 120559: contig of 7270 bp in length
* 120560 120659: gap of 100 bp
* 120660 126468: contig of 5809 bp in length
* 126469 126568: gap of 100 bp
* 126569 131619: contig of 5051 bp in length
* 131620 131719: gap of 100 bp
* 131720 136810: contig of 5091 bp in length
* 136811 136910: gap of 100 bp
* 136911 139925: contig of 3015 bp in length
* 139926 140025: gap of 100 bp
* 140026 144232: contig of 4207 bp in length
* 144233 144332: gap of 100 bp
* 144333 148902: contig of 4570 bp in length
* 148903 149002: gap of 100 bp
* 149003 152918: contig of 3316 bp in length
* 152919 153018: gap of 100 bp
* 153019 156399: contig of 3381 bp in length
* 156400 156499: gap of 100 bp
* 156500 158933: contig of 2434 bp in length
* 158934 159033: gap of 100 bp
* 159034 161884: contig of 2851 bp in length
* 161885 164301: contig of 2317 bp in length
* 164302 164401: gap of 100 bp
* 164402 166300: contig of 1899 bp in length
* 166301 166400: gap of 100 bp
* 166401 168580: contig of 2180 bp in length
* 168581 168680: gap of 100 bp
* 168681 171270: contig of 2590 bp in length
* 171271 171370: gap of 100 bp
* 171371 174092: contig of 2722 bp in length
* 174093 174192: gap of 100 bp
* 174193 176095: contig of 1903 bp in length
* 176096 176195: gap of 100 bp
* 176196 177890: contig of 1695 bp in length
* 177891 177990: gap of 100 bp
* 177991 179786: contig of 1796 bp in length
* 179787 179886: gap of 100 bp
* 179887 181478: contig of 1592 bp in length
* 181479 181578: gap of 100 bp
* 181579 182895: contig of 1317 bp in length
* 182896 182995: gap of 100 bp
* 182996 184841: contig of 1846 bp in length
* 184842 184941: gap of 100 bp
* 184942 186345: contig of 1404 bp in length
* 186346 186445: gap of 100 bp
* 186446 187635: contig of 1190 bp in length.
FEATURES
source
1. .187635
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-815L4"
1. .21332
/note="assembly_fragment"
21433. .42072
/note="assembly_fragment"
42173. .54336
/note="assembly_fragment"
54437. .66766
/note="assembly_fragment"
66867. .77849
/note="assembly_fragment"
77950. .87798
misc_feature
1. .21332
misc_feature
21433. .42072
/note="assembly_fragment"
42173. .54336
/note="assembly_fragment"
54437. .66766
/note="assembly_fragment"
66867. .77849
/note="assembly_fragment"
77950. .87798
misc_feature

misc_feature /note="assembly_fragment clone_end:T7 vector_side:right"
87899. .99510
misc_feature /note="assembly_fragment"
99611. .107273
misc_feature /note="assembly_fragment"
107374. .113189
misc_feature /note="assembly_fragment clone_end:SP6 vector_side:left"
113290. .120559
misc_feature /note="assembly_fragment"
120660. .126468
misc_feature /note="assembly_fragment"
126569. .131619
misc_feature /note="assembly_fragment"
131720. .136810
misc_feature /note="assembly_fragment"
136911. .139925
misc_feature /note="assembly_fragment"
140026. .144232
misc_feature /note="assembly_fragment"
144333. .148902
misc_feature /note="assembly_fragment"
149003. .152918
misc_feature /note="assembly_fragment"
153019. .156399

Query Match 34.4%; Score 466.2; DB 2; Length 187635;
Best Local Similarity 96.3%; Pred. NO. 1.7e-97;
Matches 488; Conservative 0; Mismatches 18; Indels 1; Gaps 1;
QY 111 TTTTACATCTACAGGTCCACCGGGCCCCAGGGGTCCACAGAGTGACAGAGGATCCCA 170
Db 68312 TTCTTTTATCTTTAGGTCCACCGGGCCCCAGGGGTCCACAGAGTGACAGAGGATCCCA 68371
QY 171 GGGACCCCTCGGCCCAACTGGCAACAGGACAGAAAGAGAGAGGGGAGCCCTGGACC 230
Db 68372 GGGACCCCTCGGCCCAACTGGCAACAGGACAGAAAGAGAGAGGGGAGCCCTGGACC 68431
QY 231 ACCTGGCCCTCGGGTGAGAGAGGCCCAATTGGACCACTGGTCCCTCCCGGAGAGCGTGG 290
Db 68432 ACCTGGCCCTCGGGTGAGAGAGGCCCAATTGGACCACTGGTCCCTCCCGGAGAGCGTGG 68491
QY 291 CGGCAAGATCTAAGGCTCCAGGGCCCCAAAGGCTCCCGTGTTCCTCCCTGGGAAGCC 350
Db 68492 CGGCAAGATCTAAGGCTCCAGGGCCCCAAAGGCTCCCGTGTTCCTCCCTGGGAAGCC 68551
QY 351 CGGCCCTCAGGCCCCAGTGGGACCCAGGCCCCCGGGCCCCACCAGGCAAGAGGGACT 410
Db 68552 CGGCCCTCA-GGCTCCAGTGGGACCCAGGCCCCCGGGCCCCACCAGGCAAGAGGGACT 68610
QY 411 CCGCGCCCTCAGGCCCCCTCTGGCTTCCAGGACTTCAGGSCACCGTTGGGGAGCCCTGG 470
Db 68611 CCGCGCCCTCAGGCCCCCTCTGGCTTCCAGGACTTCAGGSCACCGTTGGGGAGCCCTGG 68670
QY 471 GGTGCTGACCTCGGGACTGCCAGGCTTGCCTGGGTACCAAGCATGCCAGGCCCCCAA 530
Db 68671 GGTGCTGACCTCGGGACTGCCAGGCTTGCCTGGGTACCAAGCATGCCAGGCCCCCAA 68730
QY 531 GGGCCCCCGGCCCTCTCTGGCCCATCAGGAGCGTGTGCCCCGCGCCCTGCACAAATGA 590
Db 68731 GGGCCCCCGGCCCTCTCTGGCCCATCAGGAGCGTGTGCCCCGCGCCCTGCACAAATGA 68790
QY 591 GCCAACCCCGGACCGGAGCAATGG 617
Db 68791 GCCAACCCCGGACCGGAGCAATGG 68817

RESULT 14
AP000900
LOCUS AP000900 178022 bp DNA linear HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-683J11 map 18p11.3, WORKING
DRAFT SEQUENCE, 20 unordered pieces.
ACCESSION AP000900
VERSION AP000900.3 GI:8119043

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-683J11.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 178022)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,
Homo sapiens 178,022 genomic DNA of 18p11.3
Published Only in Database (1999)
2 (bases 1 to 178022)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,
Direct Submission
Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsr.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6997403.

Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsr.riken.go.jp
Project Information
Center project name: HumDraft18
Center clone name: RP11-683J11
Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 161944 bases at least Q40
Consensus quality: 170357 bases at least Q30
Consensus quality: 174322 bases at least Q20
Insert size: 176122; sum-of-contigs
Quality coverage: 4.56x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 24283 contig of 24283 bp in length
24384 46976 contig of 22492 bp in length
46976 65058 contig of 18083 bp in length
65058 81322 contig of 16164 bp in length
81322 98823 contig of 17401 bp in length
98823 110068 contig of 11145 bp in length
110068 122716 contig of 12548 bp in length
122716 133089 contig of 10273 bp in length
133089 141206 contig of 8017 bp in length
141206 149050 contig of 7744 bp in length
149050 156105 contig of 6955 bp in length
156105 162048 contig of 5843 bp in length
162048 165118 contig of 2970 bp in length
165118 167480 contig of 2262 bp in length
167480 169986 contig of 2406 bp in length
169986 171909 contig of 1823 bp in length
171909 172317 contig of 308 bp in length
172317 174663 contig of 2246 bp in length
174663 176271 contig of 1508 bp in length
176271 176372 contig of 1651 bp in length
176372 178022 contig of 178022 bp in length

Sequence updated (24-Dec-1999)
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 24283: contig of 24283 bp in length
* 24284 24383: gap of 100 bp
* 24384 46875: contig of 22492 bp in length
* 46876 46975: gap of 100 bp
* 46976 65058: contig of 18083 bp in length
* 65059 65158: gap of 100 bp
* 65159 81322: contig of 16164 bp in length
* 81323 81422: gap of 100 bp
* 81423 98823: contig of 17401 bp in length
* 98824 98923: gap of 100 bp
* 98924 110068: contig of 11145 bp in length
* 110069 110168: gap of 100 bp
* 110169 122716: contig of 12548 bp in length
* 122717 122816: gap of 100 bp
* 122817 133089: contig of 10273 bp in length
* 133090 133189: gap of 100 bp
* 133190 141206: contig of 8017 bp in length
* 141207 141306: gap of 100 bp
* 141307 149050: contig of 7744 bp in length
* 149051 149150: gap of 100 bp
* 149151 156105: contig of 6955 bp in length
* 156106 156205: gap of 100 bp
* 156206 162048: contig of 5843 bp in length
* 162049 162148: gap of 100 bp
* 162149 165118: contig of 2970 bp in length
* 165119 165218: gap of 100 bp
* 165219 167480: contig of 2262 bp in length
* 167481 167580: gap of 100 bp
* 167581 169986: contig of 2406 bp in length
* 169987 170086: gap of 100 bp
* 170087 171909: contig of 1823 bp in length
* 171910 172009: gap of 100 bp
* 172010 172317: contig of 308 bp in length
* 172318 172417: gap of 100 bp
* 172418 174663: contig of 2246 bp in length
* 174664 174763: gap of 100 bp
* 174764 176271: contig of 1508 bp in length
* 176272 176371: gap of 100 bp
* 176372 178022: contig of 1651 bp in length.

FEATURES

source
1. .178022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-683J11"
1. .24283
/note="assembly_fragment"
24384. .46875
/note="assembly_fragment"
46976. .65058
/note="assembly_fragment"
65159. .81322
/note="assembly_fragment"
81423. .98823
/note="assembly_fragment"
98924. .110068
/note="assembly_fragment"
110169. .122716
/note="assembly_fragment"
122817. .133089
/note="assembly_fragment"
133190. .141206
/note="assembly_fragment"
141307. .149050
/note="assembly_fragment"
149151. .156105
/note="assembly_fragment"
156206. .162048
/note="assembly_fragment"

misc_feature 162149. .165118
/note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature 165219. .167480
/note="assembly_fragment"
misc_feature 167581. .169986
/note="assembly_fragment"
misc_feature 170087. .171909
/note="assembly_fragment"
misc_feature 172010. .172317
/note="assembly_fragment clone_end:T7 vector_side:right"
misc_feature 172418. .174663
/note="assembly_fragment"
misc_feature 174764. .176271
/note="assembly_fragment"
misc_feature 176372. .178022
/note="assembly_fragment"
BASE COUNT 53972 a 34911 c 36303 g 50935 t 1901 others
ORIGIN

Query Match 34.4%; Score 465.8; DB 2; Length 178022;
Best Local Similarity 99.2%; Pred. No. 2e-97;
Matches 489; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
QY 126 AGGTCCACCGGGCCCCAGGGGTCCCAAGAGGTCACAGAGATCCCAAGGGACCCCTGGGCC 185
DB 123478 AGGTCCACCGGGCCCCAGGGGTCCCAAGAGGTGACAGAGATCCCAAGGGACCCCTGGGCC 123537
QY 186 AACTGGCAACAGGGACAGAAAGAGAGAGAGGGGAGCCTGGACCACCTGGGCCCTGGCGG 245
DB 123538 AACTGGCAACAGGGACAGAAAGAGAGAGAGGGGAGCCTGGACCACCTGGGCCCTGGCGG 123597
QY 246 TGAGAGAGGCCCAATTGGACCACTGGTCCCTCCCGGAGAGCCTGGCGGCAAGGATCTAA 305
DB 123598 TGAGAGAGGCCCAATTGGACCACTGGTCCCTCCCGGAGAGCCTGGCGGCAAGGATCTAA 123657
QY 306 AGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGGCC 365
DB 123658 AGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGGCC 123717
QY 366 CAGTGGGACCCAGGCCCCCGGGCCACACAGGCAAGAGGACTCCCGGGCCCTCAGGG 425
DB 123718 CAGTGGGACCCAGGCCCCCGGGCCACACAGGCAAGAGGACTCCCGGGCCCTCAGGG 123777
QY 426 CCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGAGCCTGGGGTGCCTGGACCTCG 485
DB 123778 CCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGAGCCTGGGGTGCCTGGACCTCG 123837
QY 486 GGGACTGCCAGCTTGCCTGGGGTA-CCAGGCATGCCAGGGCCCCAAGGGCCCCCGGCC 544
DB 123838 GGGACTGCCAGCTTGCCTGGGGTACCAGGCATGCCAGGGCCCCAAGGGCCCCCGGCC 123897
QY 545 CTCTGGGCCCATCAGAGCGGTGGTCCCTGGCCCTGCAGAAATGAGCAACCCCGGCAC 604
DB 123898 CTCTGGGCCCATCAGAGCGGTGGTG-CCCTGGCCCTGCAGAAATGAGCAACCCCGGCAC 123956
QY 605 CGAGGACAATGG 617
DB 123957 CGAGGACAATGG 123969

RESULT 15

AP000939/c
LOCUS AP000939 18255 bp DNA linear HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-839023 map 18p11.3, WORKING
DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION AP000939
VERSION AP000939.3 GI:8119080
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-839023.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 188255)

AUTHORS

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Tokoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 188,255 genomic DNA of 18p11.3
Published Only in DataBase (1999)
2 (bases 1 to 188255)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Tokoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)

COMMENT

On May 31, 2000 this sequence version replaced gi:6997751.

----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: Rpl1-839023
----- Summary Statistics

Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 171068 bases at least Q40
Consensus quality: 178865 bases at least Q30
Consensus quality: 182445 bases at least Q20
Insert size: 185155; sum-of-contigs
Quality coverage: 4.41x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
32 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1 16096 contig of 16096 bp in length
16197 33789 contig of 17593 bp in length
33890 51010 contig of 17121 bp in length
51111 63558 contig of 12448 bp in length
63659 75056 contig of 11398 bp in length
75157 84737 contig of 9581 bp in length
84838 90909 contig of 6072 bp in length
91010 98615 contig of 7606 bp in length
98716 105978 contig of 7263 bp in length
106079 113049 contig of 6971 bp in length
113150 119201 contig of 6052 bp in length
119302 126582 contig of 7281 bp in length
126683 132070 contig of 5388 bp in length
132171 137138 contig of 4968 bp in length
137239 143055 contig of 5817 bp in length
143156 148639 contig of 5484 bp in length
148740 153086 contig of 4347 bp in length
153187 158921 contig of 5735 bp in length
159022 163174 contig of 4153 bp in length
163275 167066 contig of 3992 bp in length
167167 170120 contig of 2954 bp in length
170121 170220 contig of 100 bp in length
172431 172430 contig of 2210 bp in length
172531 175598 contig of 3068 bp in length
175699 177118 contig of 100 bp in length
177119 177218 contig of 1420 bp in length
177219 178531 contig of 1313 bp in length
178532 178631 contig of 100 bp in length
178632 179857 contig of 1226 bp in length
179858 179957 contig of 100 bp in length
179958 181064 contig of 1107 bp in length
181065 181164 contig of 100 bp in length
181165 182735 contig of 1571 bp in length
182736 182855 contig of 100 bp in length
182856 183923 contig of 1088 bp in length
183924 184023 contig of 100 bp in length
184024 185810 contig of 1787 bp in length
185811 185910 contig of 100 bp in length
185911 187145 contig of 1235 bp in length
187146 187245 contig of 1010 bp in length
187246 188255 contig of 1010 bp in length

Sequence updated (06-Jan-2000)

FEATURES

Location/Qualifiers

Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 16096: contig of 16096 bp in length
* 16097 16196: gap of 100 bp
* 16197 33789: contig of 17593 bp in length
* 33790 33889: gap of 100 bp
* 33890 51010: contig of 17121 bp in length
* 51011 51110: gap of 100 bp
* 51111 63558: contig of 12448 bp in length
* 63559 63658: gap of 100 bp
* 63659 75056: contig of 11398 bp in length
* 75057 75156: gap of 100 bp
* 75157 84737: contig of 9581 bp in length
* 84738 84837: gap of 100 bp
* 84838 90909: contig of 6072 bp in length
* 90910 91009: gap of 100 bp
* 91010 98615: contig of 7606 bp in length
* 98616 98715: gap of 100 bp
* 98716 105978: contig of 7263 bp in length
* 105979 106078: gap of 100 bp
* 106079 113049: contig of 6971 bp in length
* 113050 113149: gap of 100 bp
* 113150 119201: contig of 6052 bp in length
* 119202 119301: gap of 100 bp
* 119302 126582: contig of 7281 bp in length
* 126583 126682: gap of 100 bp
* 126683 132070: contig of 5388 bp in length
* 132071 132170: gap of 100 bp
* 132171 137138: contig of 4968 bp in length
* 137139 137238: gap of 100 bp
* 137239 143055: contig of 5817 bp in length
* 143056 143155: gap of 100 bp
* 143156 148639: contig of 5484 bp in length
* 148640 148739: gap of 100 bp
* 148740 153086: contig of 4347 bp in length
* 153087 153186: gap of 100 bp
* 153187 158921: contig of 5735 bp in length
* 158922 159021: gap of 100 bp
* 159022 163174: contig of 4153 bp in length
* 163175 163274: gap of 100 bp
* 163275 167066: contig of 3792 bp in length
* 167067 167166: gap of 100 bp
* 167167 170120: contig of 2954 bp in length
* 170121 170220: gap of 100 bp
* 172431 172430: contig of 2210 bp in length
* 172431 172530: gap of 100 bp
* 172531 175598: contig of 3068 bp in length
* 175599 175698: gap of 100 bp
* 175699 177118: contig of 1420 bp in length
* 177119 177218: gap of 100 bp
* 177219 178531: contig of 1313 bp in length
* 178532 178631: gap of 100 bp
* 178632 179857: contig of 1226 bp in length
* 179858 179957: gap of 100 bp
* 179958 181064: contig of 1107 bp in length
* 181065 181164: gap of 100 bp
* 181165 182735: contig of 1571 bp in length
* 182736 182855: gap of 100 bp
* 182856 183923: contig of 1088 bp in length
* 183924 184023: gap of 100 bp
* 184024 185810: contig of 1787 bp in length
* 185811 185910: gap of 100 bp
* 185911 187145: contig of 1235 bp in length
* 187146 187245: gap of 100 bp
* 187246 188255: contig of 1010 bp in length.

```
source
1. .188255
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18p11.3"
/clone="RP11-839023"
1. .16096
/note="assembly_fragment"
misc_feature
16197. .33789
/note="assembly_fragment"
misc_feature
33890. .51010
/note="assembly_fragment"
misc_feature
51111. .63558
/note="assembly_fragment"
misc_feature
63659. .75056
/note="assembly_fragment"
misc_feature
75157. .84737
/note="assembly_fragment"
misc_feature
84838. .90909
/note="assembly_fragment"
misc_feature
91010. .98615
/note="assembly_fragment"
misc_feature
98716. .105978
/note="assembly_fragment"
misc_feature
106079. .113049
/note="assembly_fragment"
misc_feature
113150. .119201
/note="assembly_fragment"
misc_feature
119302. .126582
/note="assembly_fragment"
misc_feature
126683. .132070
/note="assembly_fragment"
misc_feature
132171. .137138
/note="assembly_fragment"
misc_feature
137239. .143055
/note="assembly_fragment"
misc_feature
143156. .148639
/note="assembly_fragment"
misc_feature
148740. .153086
/note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature
153187. .158921
/note="assembly_fragment"
misc_feature
159022. .163174
```

```
Query Match      32.3%; Score 437.4; DB 2; Length 188255;
Best Local Similarity 81.3%; Pred. No. 8.2e-91;
Matches 438; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 126 AGGTCACCGGGCCCCAGGGTCCAAAGAGTGACAGAGGATCCCAAGGACCCCTGGCCCC 185
Db 184461 AGGTCACCGGGCCCCAGGGTCCAAAGAGTGACAGAGGATCCCAAGGACCCCTGGCCCC 184402

QY 186 AACTGGCAACAGGAGGACAAAGGAGAGAGAGGAGGAGGAGCTGGACACCTGGCCCTGGCGG 245
Db 184401 AACTGGCAACAGGAGGAGAGAGAGGAGGAGGAGGAGGAGCTGGACACCTGGCCCTGGCGG 184342

QY 246 TGAGAGAGGCCCAATTGGACCAAGCTGGTCCCGCCGAGAGGCTGGCGGCAAGGATCTAA 305
Db 184341 TGAGAGAGGCCCAATTGGACCAAGCTGGTCCCGCCGAGAGGCTGGCGGCAAGGATCTAA 184282

QY 306 AGGCTCCAGGGCCCCCAAGGCTCCCGTGGTTCCCGCTGGGAAGCCCGCCCTCAGGGCCC 365
Db 184281 AGGCTCCAGGGCCCCCAAGGCTCCCGTGGTTCCCGCTGGGAAGCCCGCCCTCAGGGGCTC 184222

QY 366 CAGTGGGAGCCCGCCCGGGCCCCACAGGCAAGAGGAGCTCCCGGGCCCTCAGGG 425
Db 184221 CAGTGGGAGCCCGCCCGGGCCCCACAGGCAAGAGGAGCTCCCGGGCCCTCAGGG 184162

QY 426 CCTCTCTGGCTTCCAGGACTTCAGGGACCGTTGGGGAGCTGGGGTGCTGGACCTCG 485
Db 184161 CCTCTCTGGCTTCCAGGACTTCAGGGACCGTTGGGGAGCTGGGGTGCTGGACCTCG 184102

QY 486 GGGACTGCCAGGCTTGCTGGGGTACCAAGGATGCCAGGCCCAAGGGCCCCCGGGCCC 545
```

Search completed: March 21, 2003, 06:54:38
Job time : 4081.66 secs


```
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 3181 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: unknown
;   MOLECULE TYPE: cDNA
US-08-655-086-1

Query Match      8.4%; Score 113.6; DB 1; Length 3181;
Best Local Similarity 52.5%; Pred. No. 7.6e-20;
Matches 248; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY  97 CAGCTCATCAGAAATTTTACAACTACAGGTCCACCGGGCCCGCAGGGTCCAAAGAGGT 156
Db  104 CTGGTCCCAAGGCTTCCAAAGGCTTCCAAAGTCCCTCCCTGAGCCTGGCGAGCCTGGA 163

QY  157 GACAGAGGATCCAGGACCCCTGGCCCAACTGCAACAAGGGACAGAAAGGAGAGAAG 216
Db  164 GCTTCAGTCCGATGGTCCCGAGGTCCCGAGTCCCGAGTCCCTGGAAAGAATGAGATGAT 223

QY  217 GGGAGCTCTGGACCACTGGCCCTCGGGTGGAGAGAGCCCAATTTGACCACTGGTCCC 276
Db  224 GGGGAAGCTGGAACCTGGTCTGCTGAGCGTGGGCCCTCTCTGGCCCTCAGGGTCT 283

QY  277 CCCGAGAGAGCTGGCGGCAAGGATCTAAAGCTCCCGAGGGCCCGCAAGAGCTCCCGTGGT 336
Db  284 CGAGGATTCGCCGGCAACAGCTGGCCCTCGCTGGAATGAAGGGACACAGAGGTTTCAGTGGT 343

QY  337 TCCCTCGGGAAGCCCGGCCCTCAGGGCCCGCAGTGGGGACCCAGGCCCGCCGGGCCACCA 396
Db  344 TTGATGTGCCAAGGAGATGCTGGTCTGCTGCTCTAAGGGTGAAGCTGGCAGCCCT 403

QY  397 GCAAGAGGAGTCCCGGCCCTCAGGGCCCTCTGGCTTCCAGGGGACTTCAGGGGCACC 456
Db  404 GGTGAATGAGAGCTCTGTCAGATGGGCCCGCCCGTGGCCCTGCTGGTGGAGAGAGTGGC 463

QY  457 GTTGGGAGCTTGGGGTGCCTGGACCTCGGGGACTCGCAGGCTTGCCTGGGGTACCAAGGC 516
Db  464 CTGGAGGCCCTGGGCCCTGCTGGTGTGCTGGAATGATGTTGCTACTGCTGCTGCTGCGGG 523

QY  517 ATGCAGGCCCCCAGGGCCCCCGGCCCTCTCGCCCATCAGGAGCGGTGG 568
Db  524 CCCCCTGTGTCACCGGCCCGGCTGGTCTCTCTGCTTCCCTGGTGTGCTGTGG 575

RESULT 4
US-08-642-255-48
; Sequence 48, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
;   APPLICANT: CAPPELLO, Joseph
;   APPLICANT: FERRARI, Franco A.
;   TITLE OF INVENTION: High Molecular Weight Collagen-Like
;   TITLE OF INVENTION: Protein Polymers
;   NUMBER OF SEQUENCES: 135
;   CORRESPONDENCE ADDRESS:
;   ADDRESS: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
;   STREET: 4 Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/642,255
;   FILING DATE:
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: ROWLAND, Bertram I.
```

```
;
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 494-8700
;   TELEFAX: (415) 494-8771
;   TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 48:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 432 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: other nucleic acid
;   DESCRIPTION: /desc = "synthetic"
US-08-642-255-48

Query Match      7.8%; Score 106; DB 1; Length 432;
Best Local Similarity 54.0%; Pred. No. 3e-18;
Matches 217; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY  163 GGATCCCAAGGACCCCTGGCCCAACTGCAACAAGGGACAGAAAGGAGAGAGGGGAG 222
Db  1 GGTGCTCAGCGCCCAAGCAGGTCCCGAAGGGCGCATGGCCCAAGCAGGCCCGCCAAAGGTGG 60

QY  223 CCTGACCACTGGCCCTGGCGGTGAGAGAGCCCAATTGGACCACTGGTCCCCCGGGA 282
Db  61 CTGGACCGGGTGTGCTCCACCGGGTGTCTCGGGACCTGCAGGCCCGCCAGGTGCGCTCGA 120

QY  283 GAGCCTGGCGGCAAGGATCTAAAGCTCCCAAGGGCCCAAGGGTCCCGTGGTTCCTCT 342
Db  121 CGGCTGTGCTCCACCGGGTGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCTGGACCGGT 180

QY  343 GGAAGCCCGGCCCTCAGGGCCCGCAGTGGGGACCCAGGCCCGCCCGGCCCAAGGCAAA 402
Db  181 GGTCAACCGGGTGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCTGGACCGGTGGTCCA 240

QY  403 GAGGACTCCCGGCCCTCAGGGCCCTCTGGCTTCCAGGGACTTCAGGGCACTTCAGGGTGG 462
Db  241 CGGGTGTCTCGGGACCTGCAGGCCCGCCAGGTGCGCTGGACCGGTGCTCCACCGGT 300

QY  463 GAGCCTGGGGTGCCTGGACCTGGGACTGCCAGGCTTCCTGGGGTACCAAGGATGCCA 522
Db  301 GCTCCGGGACCTGCAGGCCCGCCAGGTGCGCTGCACCGGTGGTCCACCGGTGCTCCG 360

QY  523 GSCCCCAAGGGCCCCCGGCCCTCTGSCCCCATCAGGAGCG 564
Db  361 GGACCTGCAGGCCCGCCAGGTGCTCCATGGCCCAAGGAGCG 402

RESULT 5
US-09-029-348-19
; Sequence 19, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
;   APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
;   TITLE OF INVENTION: NOVEL PROCOLLAGENS
;   FILE REFERENCE: d087857PUS LISTING
;   CURRENT APPLICATION NUMBER: US/09/029,348
;   CURRENT FILING DATE: 1998-05-07
;   NUMBER OF SEQ ID NOS: 20
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 19
;   LENGTH: 1608
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
;   OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS
US-09-029-348-19

Query Match      7.8%; Score 106; DB 4; Length 1608;
Best Local Similarity 52.8%; Pred. No. 5.2e-18;
Matches 229; Conservative 0; Mismatches 205; Indels 0; Gaps 0;
```

QY 127 GGTCCACCGGGCCCCAGGGTCCAGAGGTGACAGAGGATCCAGGGACCCCTGGCCCA 186
Db 361 GGTGAACCTGGTCAAACTGGTCTGCAGTGCACCTGGTCTCATGGCCCGTGGTCT 420
QY 187 ACTGCAACAGGACACAAAGGAGAGAGGGGAGCGCTGGACCACTGGCCCTCGGGT 246
Db 421 GCTGCAACACATGAAACCGTGGTGAACCTTCTGGTCTCTTCTGGTCTCTGCTGT 480
QY 247 GAGAGAGGCCCAATTTGGACAGCTGTGCCCGCGGAGAGCGTGGCGCAAGGATCTAAA 306
Db 481 GCTGTGGCCCAAGAGTCTTAGTGGCCCAACAGGCATCTGGGCGATGAAGGAGAGCC 540
QY 307 GCTCCAGGGCCCCAAAGGCTCCCGTGTCCCTTGGGAAGCCCGCCCTCAGGGCCCC 366
Db 541 GGTGAAGAGGGCCCCAGAGGTCTTCTGCTTCAAGGGACACAATGATGCAAGGTCTG 600
QY 367 AGTGGGACCCAGGCCCCCGGGCCCCCAGGCAAGAGGAGTCCCGGGCCCTCAGGGC 426
Db 601 CCGTGTATCGTGTGTCACCATGGTGATCAAGGTGCTCTGGTGGTCTGCTGTGT 660
QY 427 CCGTGTGCTTCCAGGGACTTCAGGGCACCGTTGGGAGCCCTGGGTGCTGGACCTGG 486
Db 661 CCGTGGGCCCCGCTGCTTCTGGCCCTGCTGGAAGATGGTGGCACTGGATCT 720
QY 487 GGACGTCCAGGCTTCCCTGGGTACAGGCATGCGAGCCCAAGGGCCCCCGGGCCCT 546
Db 721 GGTAGCGTTGGACCTGCTGGCATCGAGGCCCTCAGGGTCAACCAAGGCCCTGCTG 780
QY 547 CCGTGGCCCATCAGG 560
Db 781 CCGTGGTCCCCCTGG 794

RESULT 6

US-08-494-168-1
; Sequence 1, Application US/08494168
; Patent No. 5731192
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T.
; APPLICANT: Zhou, Jing
; TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method
; TITLE OF INVENTION: of Detecting Collagen Deficiency
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/494,168
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,465
; FILING DATE: 27-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: SAXE, Bernhard D.
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 40397/104/BABR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5102 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(2..82, 86..97, 101..4399, 4403..4420, 4424
; LOCATION: ..4465, 4469..4876, 4880..5101)
US-08-494-168-1
Query Match 7.4%; Score 100; DB 1; Length 5102;
Best Local Similarity 52.4%; Pred. No. 3e-16;
Matches 220; Conservative 0; Mismatches 200; Indels 0; Gaps 0;
QY 123 ACAAGGTCCACGGGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCAGGGACCCCTGG 182
Db 2380 RAATGGTCTCGGGGAACAAGGCTTACAAGGATTACAGGGCACAAAGGATTTCTTG 2439
QY 183 CCCAACTGGCAACAAGGACACAGAAAGGAGAGAGGGGAGCCCTGGACCACTGGCCCTGC 242
Db 2440 AGACTCTGTGGCTTCCAGGACTCAAGGGTGTGCACGGGAAGCTGGCTTACTAGGCCCAA 2499
QY 243 GGGTCAGAGAGCCCCAATTTGGACCACTGTGCTCCCGGAGAGCGTGGCGCAAGGATC 302
Db 2500 AGGTGAGCGGGCACCCCTGGGACACCAAGGACAGGTGGGACAGCAGGCCACCCAGGATC 2559
QY 303 TAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTGGGAAGCCCGGCCCTCAGGG 362
Db 2560 TAGTGGTCCATATGTCATCAAGGGCAAAATCTGGGTCTCCAGGAGCACCAAGGCTTCC 2619
QY 363 CCCAGTGGGACCCAGGCCCCCGGGCCCCCAGGCAAGAGGAGTCTCCCGGCCCTCA 422
Db 2620 CATCTCAGGACATCTCTGGAAGAAAGAACAGAGGCTCTCTCTGGATCAAT 2679
QY 423 GGGCCTCTCTGGCTTCCAGGGACTTCAGGGCACCGCTGGGGAGCCTGGGTGCCTGGACC 482
Db 2680 TGTAAGAAAGGGCTGCCAGGCTTAAAGGCTTCTGGAAATCAGGCTAGTAGGACT 2739
QY 483 TCGGGACTGCCAGCTTCCCTGGGTACCAAGGCATGCGAGGCCCAAGGGCCCCCGGG 542
Db 2740 GAAAGAAAGCCAGGCTCTCCAGGGTCTGCTGGTGGTGGAGGCTCTCTGGAGCCCAAGG 2799
RESULT 7
US-07-621-091G-1
; Sequence 1, Application US/07621091G
; Patent No. 5424408
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T., Morrison, Karen E., Hudson, Billy
; APPLICANT: G.
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen
; TITLE OF INVENTION: Polynucleotides
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University, Office of Cooperative Research
; STREET: 246 Church Street
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800K storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS7.0
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/621,091G
; FILING DATE: 11/30/90
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA: No. 5424408 applicable
; ATTORNEY/AGENT INFORMATION:
; NAME: Barth, Richard S.
; REGISTRATION NUMBER: 28180
; REFERENCE/DOCKET NUMBER: 900983/RB

```

; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Calf
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
us-08-399-889-1

Query Match          7.4%; Score 99.6; DB 2; Length 1416;
Best Local Similarity 51.6%; Pred. No. 2.2e-16;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps:

QY 127 GGTCTCACCGGGCCCCAGGGGTCCAAGAGGTTGACAGAGGATCCCGAGGACCCCTCGGCCA 186
    || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 283 GGACACACAGTGTGCCCGGTACGACGAGGCCAGAGGTGATCCTGGATTCTATGGATT 342

QY 187 ACTGGCAACAGGACAGAAAGAGGAGAGAGGGGAGCCTTGGACACCTTGGCCCTTGGCGGT 246
    || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 343 CCAGGCATGAAGGAAGGGTAAATTCAGGATTTCCAGGACCCACTGGACCTCCAGGG 402

QY 247 GAGAGGCCCAATTGACACAGCTGTGTCGCCCGGAGAGCGTGGGGCAAAAGGATCTAA 306
    || || ||||| ||||| ||||| || ||||| || ||||| || ||||| || |||||
Db 403 CAAAGTGGACCAAGACACCACTGGAGTACGTGGAGAGCCTGGCACAGTGAAGATCATC 462

QY 307 GGCTCCACAGGCCCCAAAGGCTCCCGTGTTCCTCTGGGAAGCCCGCCCTCAGGGCCCC 366
    || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 463 TCCCTTCAGAGACCCAGGACCACCTGGTTACGTTGGAGAACAGGGATGCAAGGAGAA 526

QY 367 AGTGGGACCCAGGCCCCCGGGCCCCACCAGGCAAAAGAGGGACTCCCGGGCCCTCAGGGC 426
    || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 523 CCGGGCCCCAGGACCACAGAGATCCAGGACCCCTGTGGGCCAAAAGGTAACACAGG 582

QY 427 CCTCTGGCTTCAGGACATTCAGGGCACCGTGTGGGAGCGCTGGGTGCTTGGACCTCGG 486
    || | || || || || || || || || || || || || || || || || || || ||
Db 583 GAGGATGGTCCACAGGAACCTCTGTGACCAACTCGAGAAAAGGCAACAAAGGTTGTAA 642

QY 487 GGACTGCCAGGCTTGCTGGGTACCAAGGATGCCAGGCCCCCAAGGCCCCCCCCGGCCCT 546
    || || || || || || || || || || || || || || || || || || || || ||
Db 643 GGAGAGCAGGACCACTGGATCGGCTGGCGCTCCAGGCTTGAAGGGGAAACCTGGAGAC 702

QY 547 CTTGGCCCATCAGGAGCGGTGG 568
    || || || || || || || || || || || || || || || || || || || || ||
Db 703 ACTGGACCACTGCAGCAGGGG 724

RESULT 9
us-09-167-364-1
; Sequence 1, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Calf

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
US-09-167-364-1

Query Match
Best Local Similarity 7.4%; Score 99.6; DB 3; Length 1416;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 127 GTCTCCAGGGCCCGCCAGGGTCCAGAGGTGACAGAGGTCCAGGACCCCTCGGCCCA 186
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 283 GGAGCACACAGGTGTCCTCCCGTCAGCCAGGGCCAGAGGTGATCTTGATTTATGATTT 342
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 187 ACTGGCAACAAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 343 CCAGGCATGAAAGGAAGAGGTAATTCAGGATTTCCAGGACCCCTGGACCTCCAGGG 402
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 247 GAGAGGCCCAATTTGGACCACTGTCCTCCCGCCGAGAGCGTGGCGGCAAAAGATCTAAA 306
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 403 CAAAGTGGACCAAAAGAGCCACCTGGAGTACGTGGAGAGCCTGGACACAGTGAAGATCATC 462
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 307 GGCTCCAGGGCCCGCCAGGGTCCCGTGGTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 366
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 463 TCCTCTTCCAGGAAGCCAGGCCACCTGGTTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGT 522
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 367 AGTGGGGACCCAGGGCCCGCCGAGGACCCAGGACCAAGAGAGGAGTCCCGCGCCCTCAGGGC 426
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 523 CCGGGGCCCGCCAGGACCCAGGAGATCCAGGACCCCTGTGGGCCCAAAAGGTAAACAGGG 582
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 427 CCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCTGGGTGCTGGACCTCGG 486
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 583 GAGGATGGTCCACCAAGCACTCTGGACCACTGGAGAAAAGGCAACAAGGTTGTATAA 642
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 487 GGACTGCCAGGCTTGCCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCGCCCGCCCT 546
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 643 GGAGAGCAAGGACCACTGGATCCGATGCTGCCAGGCTGCCAGGCTTGAAGGGGAAACCTGGAGAC 702
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 547 CCGTGGCCCATCAGGCGGTGG 568
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 703 ACTGGACCACTGCACAGGGG 724
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 10
US-09-439-897-1
; Sequence 1, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
US-09-439-897-1

Query Match
Best Local Similarity 7.4%; Score 99.6; DB 4; Length 1416;
Matches 228; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 127 GTCTCCAGGGCCCGCCAGGGTCCAGAGGTGACAGAGGTCCAGGACCCCTCGGCCCA 186
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 283 GGAGCACACAGGTGTCCTCCCGTCAGCCAGGGCCAGAGGTGATCTTGATTTATGATTT 342
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 187 ACTGGCAACAAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246
    || |||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
```

```

; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Nucleotide-genomic DNA
; HYPOTHETICAL: No. 5691197 relevant
; ANTI-SENSE: No. 5691197 relevant
US-08-392-367B-1

Query Match          7.3%; Score 99.4; DB 1; Length 1868;
Best Local Similarity 51.5%; Pred. No. 2.8e-16;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 124 CAAGTCCACCGGGCCCGCCAGAGGTCCAAAGAGGTGACAGAGATCCACAGGACCCCTGGC 183
    ||||| || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 721 CCAGGTGTCCAAGGACCCCGCCAGGCGCCACAGGAGGAGGAGGAGGAGGAGGAGGAG 780

Qy 184 CCAACTGGCAACAAGGGACAGAAAGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 781 CTTACGGGTGCACAGGAGAAAGAGGAGCAACTGTGTCTCCAGGACCTCGAGGAGAGAG 840

Qy 244 GGTGAGAGAGGCCCAATTTGACAGCTGTGTCCTCCCGGAGAGCGTGGCGGCAAGGATCT 303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 841 GGCAGCAAGGTGACATAGGTCTCAGTGGCCCGCCAGGAGGAGGAGGAGGAGGAGGAG 900

Qy 304 AAAGCTCCAGGCGCCCAAGAGTCCCGTGTTCCTCCCTGGAGAGCGCCGCTCAGGGC 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 901 AAAGGGGACCTAGGCTTCCAGGAAACAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960

Qy 364 CCCAGTGGGACCCAGGCGCCCGCCAGGACCCAGGAGGAGGAGGAGGAGGAGGAGGAGG 423
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 961 CCCATGGGTCCTCCAGGAGTCAAGAGGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020

Qy 424 GGCCTCTCTGCTTCCAGGAGGAGTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 483
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1021 GGTTCGGTGGATCTCCAGGAGTCAAGGTGACCAAGGAGGAGGAGGAGGAGGAGGAGG 1080

Qy 484 CGGGAGTCCAGGCTTCCCTGGGTTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1081 CCAGGCGCTCAAGGTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140

Qy 544 CTCTCTGCGCCATCAGGAGGCGGTGG 568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1141 CTTCTGGGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1165

```

```

RESULT 12
US-08-893-467A-1
; Sequence 1, Application US/08893467A
; Patent No. 6063901
; GENERAL INFORMATION:
; APPLICANT: Tryggvason, Karl
; APPLICANT: Elomaa, Outi
; APPLICANT: Kangas, Maarit
; TITLE OF INVENTION: An Isolated DNA Sequence For a
; Patent No. 6063901
; TITLE OF INVENTION: No. 6063901el Macrophage Receptor with
; TITLE OF INVENTION: a Collagenous Domain and the
; TITLE OF INVENTION: Polypeptide Chain Encoded by
; TITLE OF INVENTION: such a Sequence
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fay, Sharpe, Beall, Fagan,
; ADDRESSEE: Minnich & McKee
; STREET: 1100 Superior Avenue
; STREET: Suite 700
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2518
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch,
; MEDIUM TYPE: 720 Kb storable
; COMPUTER: IBM PS/2, Model 35 SX
; OPERATING SYSTEM: DOS 5.0

```

```

; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,467A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Minnich, Richard J.
; REGISTRATION NUMBER: 24,175
; REFERENCE/DOCKET NUMBER: TRV 2 009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 861-5582
; TELEFAX: (216) 241-1666
; TELEX: (216) 980162
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1868 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Nucleotide-genomic DNA
; HYPOTHETICAL: No. 6063901 relevant
; ANTI-SENSE: No. 6063901 relevant
US-08-893-467A-1

Query Match          7.3%; Score 99.4; DB 3; Length 1868;
Best Local Similarity 51.5%; Pred. No. 2.8e-16;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 124 CAAGTCCACCGGGCCCGCCAGGAGTCCAAAGAGGTGACAGAGATCCACAGGACCCCTGGC 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 721 CCAGGTGTCCAAGGACCCCGCCAGGCGCCACAGGAGGAGGAGGAGGAGGAGGAGGAGG 780

Qy 184 CCAACTGGCAACAAGGGACAGAAAGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 781 CTTACGGGTGCACAGGAGAAAGAGGAGCAACTGTGTCTCCAGGACCTCGAGGAGAGAG 840

Qy 244 GGTGAGAGAGGCCCAATTTGACAGCTGTGTCCTCCCGGAGAGCGTGGCGGCAAGGATCT 303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 841 GGCAGCAAGGTGACATAGGTCTCAGTGGCCCGCCAGGAGGAGGAGGAGGAGGAGGAG 900

Qy 304 AAAGCTCCAGGCGCCCAAGAGTCCCGTGTTCCTCCCTGGAGAGCGCCGCTCAGGGC 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 901 AAAGGGGACCTAGGCTTCCAGGAAACAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960

Qy 364 CCCAGTGGGACCCAGGCGCCCGCCAGGACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 423
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 961 CCCATGGGTCCTCCAGGAGTCAAGAGGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020

Qy 424 GGCCTCTCTGCTTCCAGGAGGAGTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 483
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1021 GGTTCGGTGGATCTCCAGGAGTCAAGGTGACCAAGGAGGAGGAGGAGGAGGAGGAGG 1080

Qy 484 CGGGAGTCCAGGCTTCCCTGGGTTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1081 CCAGGCGCTCAAGGTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140

Qy 544 CTCTCTGCGCCATCAGGAGGCGGTGG 568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1141 CTTCTGGGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1165

RESULT 13
US-08-555-669-11
; Sequence 11, Application US/08555669
; Patent No. 5773248
; GENERAL INFORMATION:
; APPLICANT: Brewton, Richard G.
; APPLICANT: Mayne, Richard
; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas

```

; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,669
; FILING DATE: 13-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: 8389-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-854-3660
; TELEFAX: 415-854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..2098
; US-08-555-669-11

Query Match 7.3%; Score 99.4; DB 1; Length 2543;
Best Local Similarity 51.5%; Pred. No. 3.2e-16;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 124 CAAGGTCCACCGGGCCCCCAGGGGTCCAAGAGGTGACAGAGGATCCAGGGACCCCTGGC 183
DB 584 CCAGGTCCCGGACGGTGGGGTCCAGGGTTCAGAGGACCCCTGGCTACAAAGGC 643
QY 184 CCAACTGCGACACAGGACACAAAGAGAGAGAGGGGAGCCCTGGACACCTGGCCCTGGC 243
DB 644 GAGCAGGGGGAAGTCCGCAAGGACGCGGAGAGAGGTGACCTGGCCCTGGGCGGCC 703
QY 244 GGTGAGAGAGGCCCCAAAGGCTCCCGTTCCTCCCGGAGAGCGTGGCGGCAAGGATCT 303
DB 704 GGCCTCCCGGACGCTGGGGTGCAGGGCCCCCGGGATTCAGAGGACTGCCAGGGCCA 763
QY 304 AAAGGCTCCCGGACCCCAAGGCTCCCGTTCCTCCCGGAGAGCCCGGCTCAGGGC 363
DB 764 CTCGGGCCCCCTGGGACCGGGTCCCATTTGGTTCCGAGGGCCCTGGGATCCAGGA 823
QY 364 CCCAGTGGGACCCAGGCCCCCGGGCCCCACGAGCAAGAGGAGTCCCGGGCCCTCAG 423
DB 824 GCGCTGGGAAGCGGGTGAACGAGGCGAGAGGGGCCCAAGAGGTTCCGCGGCCCAAG 883
QY 424 GCGCTCTCGGCTTCAGGGACTTCAGGACACCGTTCGGGAGCCCTGGGGTGCCTGACCT 483
DB 884 GGTGACCTCGGACAGCTTGGTCCCAAGGGAACCCCGAGTGGCGGGCCCAAGCGGAG 943
QY 484 CGGGACTGCGGACCTTGCCTGGGTACCAAGGCAATCCAGGCCCCCAAGGGCCCCCGGC 543
DB 944 CCGGCATGCCAGGAAGGACGCGGCGAGATGGCGTGCCAGGACTCGATGGCCAGAAGGA 1003
QY 544 CCTCTGCGCCATCAGGACGGGTGG 568
DB 1004 GAGGCTGGTCGACAGGGTGCCTCCG 1028

RESULT 14
US-09-073-663-11
; Sequence 11, Application US/09073663

; Patent No. 6127523
; GENERAL INFORMATION:
; APPLICANT: Brewton, Richard G.
; APPLICANT: Mayne, Richard
; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McGregor & Adler, LLP
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS 8.1
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,663
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benjamin Aaron Adler, Ph.D., J.D.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5913D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 777-2321
; TELEFAX: (713) 777-6908
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..2098
; US-09-073-663-11

Query Match 7.3%; Score 99.4; DB 3; Length 2543;
Best Local Similarity 51.5%; Pred. No. 3.2e-16;
Matches 229; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 124 CAAGGTCCACCGGGCCCCCAGGGGTCCAAGAGGTGACAGAGGATCCAGGGACCCCTGGC 183
DB 584 CCAGGTCCCGGACGGTGGGGTCCAGGGTTCAGAGGACCCCTGGCTACAAAGGC 643
QY 184 CCAACTGCGACACAGGACACAAAGAGAGAGAGGGGAGCCCTGGACACCTGGCCCTGGC 243
DB 644 GAGCAGGGGGAAGTCCGCAAGGACGCGGAGAGAGGTGACCTGGCCCTGGGCGGCC 703
QY 244 GGTGAGAGAGGCCCCAAAGGCTCCCGTTCCTCCCGGAGAGCGTGGCGGCAAGGATCT 303
DB 704 GGCCTCCCGGACGCTGGGGTGCAGGGCCCCCGGGATTCAGAGGACTGCCAGGGCCA 763
QY 304 AAAGGCTCCCGGACCCCAAGGCTCCCGTTCCTCCCGGAGAGCCCGGCTCAGGGC 363
DB 764 CTCGGGCCCCCTGGGACCGGGTCCCATTTGGTTCCGAGGGCCCTGGGATCCAGGA 823
QY 364 CCCAGTGGGACCCAGGCCCCCGGGCCCCACGAGCAAGAGGAGTCCCGGGCCCTCAG 423
DB 824 GCGCTGGGAAGCGGGTGAACGAGGCGAGAGGGGCCCAAGAGGTTCCGCGGCCCAAG 883
QY 424 GCGCTCTCGGCTTCAGGGACTTCAGGACACCGTTCGGGAGCCCTGGGGTGCCTGACCT 483
DB 884 GGTGACCTCGGACAGCTTGGTCCCAAGGGAACCCCGAGTGGCGGGCCCAAGCGGAG 943
QY 484 CGGGACTGCGGACCTTGCCTGGGTACCAAGGCAATCCAGGCCCCCAAGGGCCCCCGGC 543
DB 944 CCGGCATGCCAGGAAGGACGCGGCGAGATGGCGTGCCAGGACTCGATGGCCAGAAGGA 1003

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:57:56 ; Search time 206.471 Seconds
(without alignments)

5089.177 Million cell updates/sec

Title: US-09-763-712a-1-copy_670_2024

Perfect score: 1355

Sequence: 1 atgcaacaagattgatgag.....agaaaaagggttatcatcccg 1355

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 1335.8 | 98.6 | 2930 | 10 | US-09-745-763-198 |
| c | 249 | 18.4 | 415 | 10 | Sequence 198, App |
| 3 | 111 | 8.2 | 6728 | 10 | US-09-954-456-1975 |
| 4 | 111 | 8.2 | 6728 | 10 | Sequence 1975, App |
| 5 | 110.6 | 8.2 | 1133 | 9 | US-09-954-456-782 |
| 6 | 110.6 | 8.2 | 1133 | 9 | Sequence 782, App |
| 7 | 109.4 | 8.1 | 2691 | 10 | US-09-880-107-3946 |
| 8 | 107.2 | 7.9 | 2192 | 10 | Sequence 3946, Ap |
| 9 | 107 | 7.9 | 5086 | 10 | US-09-924-340-57 |
| 10 | 107 | 7.9 | 5145 | 9 | Sequence 57, Appl |
| 11 | 107 | 7.9 | 5145 | 9 | Sequence 57, Appl |
| 12 | 107 | 7.9 | 5432 | 12 | US-09-952-600A-57 |
| 13 | 106.8 | 7.9 | 4908 | 9 | US-09-925-302-64 |
| 14 | 106 | 7.8 | 2542 | 9 | Sequence 64, Appl |
| 15 | 106 | 7.8 | 2542 | 9 | US-09-925-301-42 |
| 16 | 105.4 | 7.8 | 5416 | 10 | Sequence 42, Appl |
| 17 | 105.4 | 7.8 | 5416 | 10 | Sequence 3947, Ap |
| 18 | 105.2 | 7.8 | 6158 | 10 | Sequence 206, App |
| 19 | 105.2 | 7.8 | 6158 | 10 | US-09-925-299-206 |
| | | | | | Sequence 206, App |
| | | | | | Sequence 22, Appl |
| | | | | | Sequence 33, Appl |
| | | | | | Sequence 961, App |
| | | | | | Sequence 255, App |
| | | | | | Sequence 786, App |
| | | | | | Sequence 2094, App |
| | | | | | Sequence 6, Appl1 |
| | | | | | Sequence 762, App |

| | | | | | | |
|----|-------|-----|------|----|--------------------|-------------------|
| 20 | 104.2 | 7.7 | 4149 | 12 | US-10-044-090-104 | Sequence 104, App |
| 21 | 102.4 | 7.6 | 1486 | 10 | US-09-925-302-247 | Sequence 247, App |
| 22 | 101.6 | 7.5 | 3226 | 10 | US-09-954-456-725 | Sequence 725, App |
| 23 | 101.4 | 7.5 | 3690 | 12 | US-10-044-090-448 | Sequence 448, App |
| 24 | 98.6 | 7.3 | 2158 | 9 | US-10-001-835-98 | Sequence 98, Appl |
| 25 | 97.8 | 7.2 | 2520 | 10 | US-09-880-107-3685 | Sequence 3685, Ap |
| 26 | 95.8 | 7.1 | 1485 | 10 | US-09-925-302-246 | Sequence 246, App |
| 27 | 95.2 | 7.0 | 1797 | 9 | US-09-978-295A-613 | Sequence 613, App |
| 28 | 95.2 | 7.0 | 1797 | 9 | US-09-978-697-613 | Sequence 613, App |
| 29 | 95.2 | 7.0 | 1797 | 9 | US-09-978-192A-613 | Sequence 613, App |
| 30 | 95.2 | 7.0 | 1797 | 9 | US-09-998-832A-613 | Sequence 613, App |
| 31 | 95.2 | 7.0 | 1797 | 9 | US-09-978-189-613 | Sequence 613, App |
| 32 | 95.2 | 7.0 | 1797 | 9 | US-10-174-590-331 | Sequence 331, App |
| 33 | 95.2 | 7.0 | 1797 | 9 | US-10-176-758-331 | Sequence 331, App |
| 34 | 95.2 | 7.0 | 1797 | 9 | US-10-175-737-331 | Sequence 331, App |
| 35 | 95.2 | 7.0 | 1797 | 9 | US-10-173-706-331 | Sequence 331, App |
| 36 | 95.2 | 7.0 | 1797 | 9 | US-10-175-738-331 | Sequence 331, App |
| 37 | 95.2 | 7.0 | 1797 | 9 | US-10-175-752-331 | Sequence 331, App |
| 38 | 95.2 | 7.0 | 1797 | 9 | US-10-176-482-331 | Sequence 331, App |
| 39 | 95.2 | 7.0 | 1797 | 9 | US-10-176-757-331 | Sequence 331, App |
| 40 | 95.2 | 7.0 | 1797 | 9 | US-10-176-913-331 | Sequence 331, App |
| 41 | 95.2 | 7.0 | 1797 | 9 | US-10-180-552-331 | Sequence 331, App |
| 42 | 95.2 | 7.0 | 1797 | 9 | US-10-180-557-331 | Sequence 331, App |
| 43 | 95.2 | 7.0 | 1797 | 9 | US-10-173-700-331 | Sequence 331, App |
| 44 | 95.2 | 7.0 | 1797 | 9 | US-10-174-572-331 | Sequence 331, App |
| 45 | 95.2 | 7.0 | 1797 | 9 | US-10-174-579-331 | Sequence 331, App |

ALIGNMENTS

RESULT 1

US-09-745-763-198
; Sequence 198, Application US/09745763
; Patent No. US20020065394A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth
McCoy, John M.
LaVallie, Edward R.
Collins-Racie, Lisa A.
Evans, Cheryl
Merberg, David
Treacy, Maurice
Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/745,763
FILING DATE: 18-Jun-2000
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 198:
SEQUENCE CHARACTERISTICS:
LENGTH: 2930 base pairs
TYPE: nucleic acid

| | | | | | | | | | |
|--|------|---|------|--|--|--|--|--|--|
| ; STRANDEDNESS: double | | | | | | | | | |
| ; TOPOLOGY: linear | | | | | | | | | |
| ; MOLECULE TYPE: cDNA | | | | | | | | | |
| ; SEQUENCE DESCRIPTION: SEQ ID NO: 198: | | | | | | | | | |
| US-09-745-763-198 | | | | | | | | | |
| Query Match | | | | | | | | | |
| Best Local Similarity 98.6%; Score 1335.8; DB 10; Length 2930; | | | | | | | | | |
| Matches 1348; Conservative 0; Mismatches 2; Indels 1; Gaps 1; | | | | | | | | | |
| QY | 1 | ATGCAACAGATTTCATGAGTTCGAGGTAGACACTGAAGTAGCCAACTTATCAGTGATT | 60 | | | | | | |
| DB | 1266 | ATGCAACAAGATTTCATGAGTTCGAGGTTCAGACTGAAGTAGCCAACTTATCAGTGATT | 1325 | | | | | | |
| QY | 61 | ATGGAAGAAATGAAGTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAATTTTACAATA | 120 | | | | | | |
| DB | 1326 | ATGGAAGAAATGAAGTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAAATTTTACAATA | 1385 | | | | | | |
| QY | 121 | CTAAGAGTCCACCGGGCCCCAGGGTCCAAGAGTGCAGAGGATCCAGGAGCCCCCT | 180 | | | | | | |
| DB | 1386 | CTAAGAGTCCACCGGGCCCCAGGGTCCAAGAGTGCAGAGGATCCAGGAGCCCCCT | 1445 | | | | | | |
| QY | 181 | GGCCCAACTGGCAACAAGGACAGAAAGGAGAGAGAGGGGAGCCTGGACCACTGGCCCT | 240 | | | | | | |
| DB | 1446 | GGCCCAACTGGCAACAAGGACAGAAAGGAGAGAGGGGAGCCTGGACCACTGGCCCT | 1505 | | | | | | |
| QY | 241 | GGGGTGTAGAGAGGCCCAATTTGGACAGCTGGTCCCGCCGAGAGCGTGGCGCAAGGA | 300 | | | | | | |
| DB | 1506 | GGGGTGTAGAGAGGCCCAATTTGGACAGCTGGTCCCGCCGAGAGCGTGGCGCAAGGA | 1565 | | | | | | |
| QY | 301 | TCTAAGGCTCCAGGGCCCCAAGGCTCCCGTGGTTCCCTTGGGAAGCCCGGCCCTCAG | 360 | | | | | | |
| DB | 1566 | TCTAAGGCTCCAGGGCCCCAAGGCTCCCGTGGTTCCCTTGGGAAGCCCGGCCCTCAG | 1625 | | | | | | |
| QY | 361 | GGCCCACTGGGACCCAGGCCCCCGGGCCCAAGGAGGAGGAGGACTCCCGGCCCT | 420 | | | | | | |
| DB | 1626 | GGCCCACTGGGACCCAGGCCCCCGGGCCCAAGGAGGAGGAGGACTCCCGGCCCT | 1685 | | | | | | |
| QY | 421 | CAGGGCCCTCTCGGTTCCAGGGACTTCCAGGCACCGTTGGGGAGCCTGGGGTGGCTGA | 480 | | | | | | |
| DB | 1686 | CAGGGCCCTCTCGGTTCCAGGGACTTCCAGGCACCGTTGGGGAGCCTGGGGTGGCTGA | 1745 | | | | | | |
| QY | 481 | CCTCGGGGACTCCAGGCTTGGCTGGGTACCAAGCATGCCAGGCCCCCAAGGGCCCCCG | 540 | | | | | | |
| DB | 1746 | CCTCGGGGACTCCAGGCTTGGCTGGGTACCAAGCATGCCAGGCCCCCAAGGGCCCCCG | 1805 | | | | | | |
| QY | 541 | GGCCCTCTCTGGCCCATCAGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCCG | 600 | | | | | | |
| DB | 1806 | GGCCCTCTCTGGCCCATCAGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCCG | 1865 | | | | | | |
| QY | 601 | GCACGGAGGACAATGGCTGCGCCCTCAGTGGAAAGCTTCACAGACAAATGCTACTAT | 660 | | | | | | |
| DB | 1866 | GCACGGAGGACAATGGCTGCGCCCTCAGTGGAAAGCTTCACAGACAAATGCTACTAT | 1925 | | | | | | |
| QY | 661 | TTTTTCAGTTGAAAAGAAATTTTCAGGATGCAAGCTTTCTGTGAAGACAAGTCTTCA | 720 | | | | | | |
| DB | 1926 | TTTTTCAGTTGAAAAGAAATTTTCAGGATGCAAGCTTTCTGTGAAGACAAGTCTTCA | 1985 | | | | | | |
| QY | 721 | CATCTTGTGTTTTCATAAAGACTAGAGAGGAACAGCAATGGATAAAAAACAGATGTTAGG | 780 | | | | | | |
| DB | 1986 | CATCTTGTGTTTTCATAAAGACTAGAGAGGAACAGCAATGGATAAAAAACAGATGTTAGG | 2045 | | | | | | |
| QY | 781 | AGAGAGGCCACTGGATCGGCTCAGACTCAGACGCTGAAATGAATGGAAGTGGGTG | 840 | | | | | | |
| DB | 2046 | AGAGAGGCCACTGGATCGGCTCAGACTCAGACGCTGAAATGAATGGAAGTGGGTG | 2105 | | | | | | |
| QY | 841 | GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTCAT | 900 | | | | | | |
| DB | 2106 | GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGGTCAT | 2165 | | | | | | |
| QY | 901 | GGCCATGGGCCAGGAGAGACTGTCTGGTTGATTATGCTGGCAGTGGAAACCATTTTC | 960 | | | | | | |
| DB | 2166 | GGCCATGGGCCAGGAGAGACTGTCTGGTTGATTATGCTGGCAGTGGAAACCATTTTC | 2225 | | | | | | |

| | | | |
|----|------|---|------|
| QY | 961 | CAATGTGAAGACGTCAATAAATTCATTTCCGAAAAAGACAGGAGACAGTACTGTCTCT | 1020 |
| DB | 2226 | CAATGTGAAGACGTCAATAAATTCATTTCCGAAAAAGACAGGAGACAGTACTGTCTCT | 2285 |
| QY | 1021 | GCATTATAACGGACTGTGATGGGATCAGATGACCAAAATTTTCAGCTCTCAAAGGCAAGG | 1080 |
| DB | 2286 | GCATTATAACGGACTGTGATGGGATCAGATGACCAAAATTTTCAGCTCTCAAAGGCAAGG | 2345 |
| QY | 1081 | ACACTCCCTTTCTAATTTGCATCACTTCTCATCAGATTGAAAAAAGAAAGCACTCAAAA | 1140 |
| DB | 2346 | ACACTCCCTTTCTAATTTGCATCACTTCTCATCAGATTGAAAAAAGAAAGCACTCAAAA | 2404 |
| QY | 1141 | CCAAATTACTGAAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCCGTCATTACCCAAAGAC | 1200 |
| DB | 2405 | CCAAATTACTGAAAAAAATTTGACAGCTAGTGTGTTTTTACCATCCCGTCATTACCCAAAGAC | 2464 |
| QY | 1201 | TTGGGAACATAAATGTTCCCAAGGTGATATGCTGATTTTCATTGTGCACATGGACTGAA | 1260 |
| DB | 2465 | TTGGGAACATAAATGTTCCCAAGGTGATATGCTGATTTTCATTGTGCACATGGACTGAA | 2524 |
| QY | 1261 | TCACATAGATTCTCTCCGTCCGTACGTACCGTGGGATATACAAATATGCTTTCCAAAGTA | 1320 |
| DB | 2525 | TCACATAGATTCTCTCCGTCCGTACGTACCGTGGGATATACAAATATGCTTTCCAAAGTA | 2584 |
| QY | 1321 | TGGAACACTTCCAATCAGAAAAAGGTTATCAT | 1351 |
| DB | 2585 | TGGAACACTTCCAATCAGAAAAAGGTTATCAT | 2615 |

RESULT 2

US-09-954-456-1975/c
; Sequence 1975, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE OF INVENTION: Sets
; CURRENT FILING DATE: 2001-09-18
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1975
; LENGTH: 415
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: n=a,t,g or c
US-09-954-456-1975

Query Match 18.4%; Score 249; DB 10; Length 415;

Best Local Similarity 95.3%; Pred. No. 1.3e-55;
Matches 266; Conservative 0; Mismatches 12; Indels 1; Gaps 1;
Qy 1073 GCAAGGACACCTCTTCTTAATTCATCATCACCTCTCATAGATTGAAAAAAGC 1132
Db 415 GCAAGGACACACCTCTTCTTAATTCATCATCACCTCTCATAGATTG-AAAAAAGC 357
Qy 1133 ACTGAAACCAATTAATGAAAAAATTCAGAGCTAGTGTGTTTACCATCGTCATAC 1192
Db 356 ACTGAAACCAATTAATGAAAAAATTCAGAGCTAGTGTGTTTACCATCGTCATAC 297
Qy 1193 CCAAGACTTGGGACTAAATGTTCCCGAGGCTATGCTGATTTTCATTTGTCACAT 1252
Db 236 CCAAGACTTGGGACTAAATGTTCCCGAGGCTATGCTGATTTTCATTTGTCACAT 237
Qy 1253 GGACTGAATCATAGATTCTCTCCGTCAGTAAACCGGCGATTATACAAATTATGCTT 1312
Db 236 GGACTGAATCATAGATTCTCTCCGTCAGTAAACCGGCGATTATACAAATTATGCTT 177
Qy 1313 CCAAGTATGGAACACTCCAATCAGAAAAAGTTATCAT 1351
Db 176 CCAAGTATGGAACACTCCAATCAGAAAAAGTTATCAT 138

RESULT 3

US-09-954-456-782
; Sequence 782, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc

; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 782

LENGTH: 6728

TYPE: DNA

ORGANISM: Homo sapiens

US-09-954-456-782

Query Match 8.2%; Score 111; DB 10; Length 6728;
Best Local Similarity 53.0%; Pred. No. 9.1e-19;
Matches 237; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

Qy 122 TACAAGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGATCCACGAGACCCCTG 181
Db 721 TCCAAGTCCCTCTGTGAGCTGCGAGCTTGAGCTTCAGGTCCATGGTCCCCGAG 780
Qy 182 GCCCAACTGGCAACAGGACAGAAAGGAGAGAGGGGAGCTGACACCTGCGCCCTG 241

Db 781 GTCCCCAGGTCCCCCTCGAAAGAAATGGAGATGATGGGAAGCTGGAAAAACCTGTCGTC 840
Qy 242 CGGGTGAGAGAGGCCCAATTTGACCAGCTGGTCCCGCCGAGAGCGTGGCGCAAGGAT 301
Db 841 CTGGTGAGCGTGGGCTCTCTGGCCCTCAGGCTGCTCGAGGATTCGCCGAAACAGCTGGCC 900
Qy 302 CTAAGAGTCCACAGGCCCCCAAGAGCTCCGCTGGTTCCTCCCTGGGAAGCCCGCCCTCAGG 361
Db 901 TCCCTGGAATGAAGGACACAGAGGTTTCAGTGGTTTGGATGGTGCCAAAGGAGATGCTG 960
Qy 362 GCCCAGTGGGACCCAGGCCCCCGGGCCACACAGGACCAAGAGGACTCCCGCGCCCTC 421
Db 961 GTCTGCTGGTCTTAAGGCTGAGCCTGGCAGCCCTGGTGAATGGAGCTCCTGTGTCAGA 1020
Qy 422 AGGGCCCTCTCTGGCTTCAGGGACTTCAGGGCACCGTTGAGGAGCTGGGGGCTGGCTGGAC 481
Db 1021 TGGGCCCCGCTGGCTGCTGCTGAGAGAGTGGCCCTGGAGCCCTGCGCCCTGCTGGTG 1080
Qy 482 CTCGGGAGTCCAGGCTTGGCTGGGTTACCAAGGATGCGAGGCCCCCAAGGCCCCCCCCG 541
Db 1081 CTCGTGGAATGATGCTGCTACTGCTGCTGCGCGGCCCTGTCCTCCACCGCCCGCTG 1140
Qy 542 GCCCTCTGGCCATCAGGAGCGGTGG 568
Db 1141 GTCTCTCTGGCTTCCCTGCTGCTGTTG 1167

RESULT 4

US-09-880-107-3946
; Sequence 3946, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3946
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274615
US-09-880-107-3946

Query Match 8.2%; Score 111; DB 10; Length 6728;
Best Local Similarity 53.0%; Pred. No. 9.1e-19;
Matches 237; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

Qy 122 TACAAGTCCACCGGGCCCCAGGGTCCCAAGAGGTGACAGAGATCCACGAGACCCCTG 181
Db 721 TCCAAGTCCCTCTGTGAGCTTGGCGAGCTTCAGTTCCTGCTGGTCCCGGAG 780
Qy 182 GCCCAACTGGCAACAGGACAGAAAGGAGAGAGGGGAGCTGACACCTGCGCCCTG 241
Db 781 GTCCCCAGTCCCTCGAAAGAAATGGAGATGATGGGAAGCTGGAAACCTGTCGTC 840
Qy 242 CGGGTGAGAGAGGCCCAATTTGACCAGCTGGTCCCGCCGAGAGCGTGGCGCAAGGAT 301
Db 841 CTGGTGAGCGTGGGCTCTCTGGCCCTCAGGCTGCTCGAGGATTCGCCGAAACAGCTGGCC 900
Qy 302 CTAAGAGTCCACGAGGCCCCCAAGGCTCCGCTGGTTCCTCCCTGGGAAGCCCGCCCTCAGG 361
Db 901 TCCCTGGAATGAAGGACACAGAGGTTTCAGTGGTTTGGATGGTGCCAAAGGAGATGCTG 960

| | | | |
|----|------|---|------|
| Qy | 362 | GCCCCAGTGGGACCCAGGCCCCCAGGCCACCAGGCANAAGAGGGACTCCCCGGGCCCTC | 421 |
| | | | |
| Db | 961 | GTCTCTGTCGTCTTAAGGCTGAGCGCTTGCAAGCTGTTGAAAATGGAGCTCTCTGGTCAGA | 1020 |
| | | | |
| Qy | 422 | AGGGCCCTCCTGGCTTCAGGGAACCTTAGAGGCACCGTTGGGGAGCCTTGGGTCCTGGAC | 481 |
| | | | |
| Db | 1021 | TGGGCCCCGCTGGCTTCGCTGTGTGAGAGAGGTTCGCGCTTGGAGCCCTTGGCCCTGCTGGTG | 1080 |
| | | | |
| Qy | 482 | CTCGGGGACTCGCAGGCTTGCCTGGGTACCAAGGCATGCCAGGCCCCAAAGGCCCCGCCG | 541 |
| | | | |
| Db | 1081 | CTCGTGGAAATGATGCTGCTACTGCTGTGTCGGGCCCTCTGGTCCACCGGCCCCGCTG | 1140 |
| | | | |
| Qy | 542 | GCCTCTCTGGCCCATCAGGAGCGGTGG | 568 |
| | | | |
| Db | 1141 | GTCTCTGGCTTCCCTGGTCTGTG | 1167 |
| | | | |

RESULT 5
US-09-924-340-57
; Sequence 57, Application US/09924340
; Publication No. US20030027248A1
; GENERAL INFORMATION:
; APPLICANT: Bejanin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US2.REG
; CURRENT APPLICATION NUMBER: US/09/924,340
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: Jpatent
; SEQ ID NO 57

Query Match 8.2%; Score 110.6; DB 9; Length 1133;
Best Local Similarity 52.9%; Pred. NO. 4.4e-19;
Matches 267; Conservative 0; Mismatches 229; Indels 9; Gaps 1;

| | | | |
|----|-----|--|-----|
| Qy | 90 | GCATGTGACGCTCATCAAGAAATTTTACAATACTACAAGTCCACCGGGCCCAAGGGGTCC | 149 |
| | | | |
| Db | 109 | GCAAGACAGGGCCCCCAGGATCTCAAGGCCTCCCAAGCCGCCCAAGGCCACAGCGGTCC | 168 |
| | | | |
| Qy | 150 | AAGAGGTGACAGGATCCAGGAGCCCCCTGGCCCACTGGCAACAAGGGACACAAGG | 209 |
| | | | |
| Db | 169 | CAGAGGAGCGAGAGCCCCAAGGTAACTCCGGTGAAGAGGCGCACAGGGGATTTCAAGG | 228 |
| | | | |
| Qy | 210 | AGAGAGGGGGAGCCTTGACCACCTCGGCCTCGGGTGAAGAGGCCCAATTGGACCACG | 269 |
| | | | |
| Db | 229 | CCAGCAGAGGCTTTTCGGGGGCCACCGGTCCTCGATTCCCAAGCAAGTGTGATCAC | 288 |
| | | | |
| Qy | 270 | TGTCCTCCCGG-----AGAGCGTGGCGCAAGGATCTAAAGGCTCCCAAGGGCC | 320 |

| | | | |
|----|-----|--|-----|
| Db | 289 | TGCGCCACCTGGGCCCTCAAGCAGAGAAGGACGACGAAGGATTCGAGGCCCCATCAGGCCT | 348 |
| Qy | 321 | CAAAGGCTCCCGTGTTCCTCGGGAAGCCCGGCCCTCAGGGCCCCAGGTGGGACCCAGG | 380 |
| Db | 349 | GCCTGGCTCCCTGGGCCACCGGAGCTCTCTGGGATTCAGGGCCCCGCGGTCTGGATGG | 408 |
| Qy | 381 | CCCCCGGGCCCCACGAGGAAAGAGGACTCCCGGCCCTCAGGGCCCCCTCTCTGGCTTCCA | 440 |
| Db | 409 | TTTGGATGGGAAGGATGGCAAGCCTGGCTTGAGGGGGACCCCTGGTCTGCTGGCCCCC | 468 |
| Qy | 441 | GGACTTTCAGGGCACCGTTTGGGAGAGCCTGGGGTGCCTTGGACCTTCGGGGACTGCACGGCTT | 500 |
| Db | 469 | TGGACTCATGGGACCAACCGGGCTTTAAGGGGAAACAGGACATCTCTGGCTCTCCAGGACC | 528 |
| Qy | 501 | GCCTGGGGTACCAAGGATGCCAGGCCCAAGGGCCCCCGCGCCCTCTCTGGCCCCATCAGG | 560 |
| Db | 529 | TAGGGTGTACTGTGGCAACACAGGTCTCTCTGGCAGCATGGCCCGCCCTGGCGCAGAGGG | 588 |
| Qy | 561 | AGCGGTGGTCCCTGTGSCCTTCGAC | 585 |
| Db | 589 | TGAACCTTGGTGCATGGGACCCACAG | 613 |

RESULT 6
US-09-992-600A-57
; Sequence 57, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjanin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN DNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US4.DIV
; CURRENT APPLICATION NUMBER: US/09/992.600A

```

? CURENT FILING DATE: 2001-11-13
? PRIOR APPLICATION NUMBER: US 09/934,340
? PRIOR FILING DATE: 2001-08-06
? PRIOR APPLICATION NUMBER: PCT/IB01/01715
? PRIOR FILING DATE: 2001-08-06
? PRIOR APPLICATION NUMBER: US 60/305,456
? PRIOR FILING DATE: 2001-07-13
? PRIOR APPLICATION NUMBER: US 60/302,277
? PRIOR FILING DATE: 2001-06-29
? PRIOR APPLICATION NUMBER: US 60/298,698
? PRIOR FILING DATE: 2001-06-15
? PRIOR APPLICATION NUMBER: US 60/293,574
? PRIOR FILING DATE: 2001-05-25
? NUMBER OF SEQ ID NOS: 114
? SOFTWARE: Jpatent
? SEQ ID NO 57

```

```

1 LENGTH: 1133
2 TYPE: DNA
3 ORGANISM: Homo sapiens
4 FEATURE:
5 NAME/KEY: 5'UTR
6 LOCATION: 1..475
7 FEATURE:
8 NAME/KEY: CDS
9 LOCATION: 476..964
10 FEATURE:
11 NAME/KEY: 3'UTR
12 LOCATION: 965..1133
13 FEATURE:
14 NAME/KEY: polyA_signal
15 LOCATION: 1101..1106
16 FEATURE:
17 NAME/KEY: polyA_site
18 LOCATION: 1118..1133
19 US-09-993-600A-57

```

Query Match 8.2%; Score 110.6; DB 9; Length 1133;
Best Local Similarity 52.9%; Pred. No. 4.4e-19;
Matches 267; Conservative 0; Mismatches 229; Indels 9;

Db 818 GGGACACAGGACAGAGGATCAAGAGGTACCGTGGTCTCTGGCTCCAGGCTCC- 876
QY 367 AGTGGGACCCAGGCCCCCGGCCCCCAGGCAAAAGAGGAGTCCCGGCCCTCAGGCG 426
Db 877 -----CCCTGGCCCTCCTGGCTCTCTGGTGAACAGGTCCCTCTGGAGCCTCTGGT 928
QY 427 CCTCTGGCTTCCAGGACTTCAGGACACCGTTGGGAGGACCTGGGTGCTGACACTCGG 486
Db 929 CTTGCTGGTCCCGAGGTCCTGGCTCTGGTGGTCTGGTGGTCTGGTGGTCTGGTGGT 988
QY 487 GGAGTCCAGGCTTGGCTGGGTACCGATGCCAGGCCCCCAAGGCCCCCGGCCCT 546
Db 989 GGTCTCCCTGGGCCCCATTTGGGCCCCCTGGTCTGGTGGTGGTGGTGGTGGTGGTGGT 1048
QY 547 CTGGGCCCATCAGGAGCGG-TGGTGGCCCTGGCCCTGCAGAAATGAGCCAAACCCGCGCACC 605
Db 1049 GTTGGTCCCCCGGCCCTCCTGGACCTCCTGGTCCCTGGTCCCTGGTCCCTGGTCCCTGGTCC 1108
QY 606 GGAGGACAATGGTGGCCGCCCTCACTGGAAGA 637
Db 1109 GACTTCAGCTTCTGCCCCCAGCCACCTCAAGA 1140

RESULT 9
US-09-880-107-3947
; Sequence 3947, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherif, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3947
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274616
; NAME/KEY: unsure
; LOCATION: (1)..(5086)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3947

Query Match 7.9%; Score 107; DB 10; Length 5086;
Best Local Similarity 51.3%; Pred. No. 8.7e-18;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

QY 120 ACTACAAGTTCACCGGGCCCCAGGGTCCAAAGAGTGACAGAGTCCACGGAGCCCC 179
Db 709 ACAGCCCGTGTCTCTGGTGTGAAGGTGAACCTGGTGGTCCCTGGTGAAGTGAAGTCC 768
QY 180 TGGCCCACTGCTCAACAGGACAGAAAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGG 239
Db 769 AGGTCAACAGGAGGCGCGTGGCTCTCTGGTGAAGAGACGTTGGTGGTGGTGGTGGTGG 828
QY 240 TGGGGGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 299
Db 829 AGCTGGTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 888
QY 300 ATCTAAAGGCTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 359
Db 889 GTCTGCTGGCCCTCCAGGCTTCCAGGCTCCCGTGGCCCGCCAGGAGGAGGAGGAGGAGG 948

QY 360 GGGCCCCAAGTGGGAGACCCAGGCCCCCGGCCCCCAGGCAAAAGAGGAGTCCCGGCCCC 419
Db 949 TGGTAACGCTGGTCCCTGGTGGTCCCGCGGTCCTGGTGAAGTGGTGGTGGTGGTGGTGG 1008
QY 420 TCAGGGCCCTCCTGGCTTCCAGGAGCTTCAGGACACCGTTGGGAGCCTTGGGAGCCTGG 479
Db 1009 CTCGGGCCCCGTTGGACCTCCTGGTAATCCTGGAGCAACAGCGCTTACTGGTGCCAAGG 1068
QY 480 ACTCGGGAGCTGCCAGGCTTGGCTGGGTACCGAGCATGCCAGGCCCCCAAGGCCCCCCC 539
Db 1069 TGTCTGGCTTCCCGGGCTTGGTGGGCTCCCGGCTCCCTGGAGCCCCCGGCTATTCC 1128
QY 540 CGGCCCTCCTCGGCCCATCAGGAGCGGTGGTGGCCCTTGGCCCTTGCAGAGTGAAGCAACCC 599
Db 1129 TGGCCCTGTGGTGGTGGCGGTGCTACTGGTGCAGAGACTTGTGGTGAGCCTGGTCC 1188
QY 600 GGC 602
Db 1189 AGC 1191

RESULT 10
US-09-925-299-206
; Sequence 206, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Query Match 7.9%; Score 107; DB 9; Length 5145;
Best Local Similarity 51.3%; Pred. No. 8.8e-18;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

QY 120 ACTACAAGTTCACCGGGCCCCAGGGTCCAAAGAGTGACAGAGTCCACGGAGCCCC 179
Db 712 ACAGCCCGTGTCTCTGGTGTGAAGGTCAACCTGGTGGTGGTGGTGGTGGTGGTGGTGG 771
QY 180 TGGCCCACTGCTCAACAGGAGCAGACAGAAAGAGAGAGGAGGAGCCTGGACCACCTGGCCC 239
Db 772 AGGTCAACAGGAGGCGCGTGGGCTTCTGGTGAAGAGACGTTGGTGGTGGTGGTGGTGG 831
QY 240 TCGGGGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 299
Db 832 AGCTGGTGGCCCTGGCAGTGAATGGAAGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 891

Qy 300 ATCTAAGGCTCCAGGCCCCCAAGAGCTCCCGTGGTTCCCTGGGAAGCCCGGGCCCTCA 359
Db 892 GTCTGCTGGCCCTCCAGGCTTCCAGAGTCCCTGGCCCAAGGCTGAAATGGARCTGT 951
Qy 360 GGGCCCCAGTGGGACCCAGGCCCCCGGGCCCCACCAAGGAGAGGAGCTCCCGGGCCC 419
Db 952 TGGTAACGCTGGTCTGCTGGTCCCGCGGTCCTGGTGAAGTGGGTCTTCCAGGCT 1011
Qy 420 TCAGGCCCCCTCTGCTTCCAGGACCTTCAGGGCACCTTGGGGAGCCCTGGGGTGCCTGG 479
Db 1012 CTCGGCCCCGCTGAGACCTCTGCTTAATCCTGGAGCAACGSCCTACTGGTGGCAAGG 1071
Qy 480 ACCTCGGGAGCTGACAGCTTCCCTGGGTACAGGGACCTTGGGGAGCCCTGGGGTGCCTGG 479
Db 1072 TGTGCTGGGCTTCCCGGCTGCTGGGGCTCCCGGCTCCCTGGAGCCCGGCTATTCC 1131
Qy 540 CGGCCCTCTGCCCCATCAGGAGCGGTGGTCCCTGCGCTGCAGATGAGCCCAACCCC 599
Db 1132 TGGCCCTGTTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1191
Qy 600 GGC 602
Db 1192 AGC 1194

RESULT 11

US-09-925-299-206
; Sequence 206, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR FILING DATE: 2000-03-08
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Query Match 7.9%; Score 107; DB 10; Length 5145;
Best Local Similarity 51.3%; Pred. No. 8.8e-18;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

Qy 120 ACTACAGGTCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCCAAGGACCCCC 179
Db 712 ACAGCCGGTCTCCTGTTGAAGGTGAACCTGGTCCCTCGTGAAGATGGAACCTCC 771
Qy 180 TGGCCCACTGGCAACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 239
Db 772 AGGTCAACAG 831
Qy 240 TSCGGGTGAG 299

Db 832 AGCTGGTCCCGTGGCAGTGTGAAGTGTGGGTCCCGTGGGTCTGCTGTGCTCCATTGG 891
Qy 300 ATCTAAGGCTCCAGGCCCCCAAGAGCTCCCGTGGTTCCCTGGGAAGCCCGGGCCCTCA 359
Db 892 GTCTGCTGGCCCTCCAGGCTTCCAGAGTCCCTGGCCCAAGGCTGAAATGGARCTGT 951
Qy 360 GGGCCCCAGTGGGACCCAGGCCCCCGGGCCCCACCAAGGAGAGGAGCTCCCGGGCCC 419
Db 952 TGGTAACGCTGGTCTGCTGGTCCCGCGGTCCTGGTGAAGTGGGTCTTCCAGGCT 1011
Qy 420 TCAGGCCCCCTCTGCTTCCAGGACCTTCAGGGCACCTTGGGGAGCCCTGGGGTGCCTGG 479
Db 1012 CTCGGCCCCGCTGAGACCTCTGCTTAATCCTGGAGCAACGSCCTACTGGTGGCAAGG 1071
Qy 480 ACCTCGGGAGCTGACAGCTTCCCTGGGTACAGGGACCTTGGGGAGCCCTGGGGTGCCTGG 479
Db 1072 TGTGCTGGGCTTCCCGGCTGCTGGGGCTCCCGGCTCCCTGGAGCCCGGCTATTCC 1131
Qy 540 CGGCCCTCTGCCCCATCAGGAGCGGTGGTCCCTGCGCTGCAGATGAGCCCAACCCC 599
Db 1132 TGGCCCTGTTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1191
Qy 600 GGC 602
Db 1192 AGC 1194

RESULT 12

US-10-044-090-22
; Sequence 22, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 5432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.13
US-10-044-090-22

Query Match 7.9%; Score 107; DB 12; Length 5432;
Best Local Similarity 51.3%; Pred. No. 9e-18;
Matches 248; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

Qy 120 ACTACAGGTCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCCAAGGACCCCC 179
Db 1042 ACAGCCGGTCTCCTGTTGAAGGTGAACCTGGTCCCTCGTGAAGATGGAACCTCC 1101
Qy 180 TGGCCCACTGGCAACAGGAGACAGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 239
Db 1102 AGGTCAACAG 1161
Qy 240 TCGGGGTGAG 299
Db 1162 AGCTGGTCCCGTGGCAGTGTGAAGTGTGGTCCCGTGGGTCTGCTGTGCTCCATTGG 1221
Qy 300 ATCTAAGGCTCCAGGCCCCCAAGAGCTCCCGTGGTTCCCTGGGAAGCCCGGGCCCTCA 359
Db 1222 GTCTGCTGGCTCCAGGCTTCCAGAGTCCCTGGCCCAAGGCTGAAATGGAGCTGT 1281
Qy 360 GGGCCCCAGTGGGAG 419
Db 1282 TGGTAACGCTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1341


```
QY 542 GCCCTCCTGGCCCATCAGAGCGGTGGTGCCTC 574
    ||| ||||| | ||||| | |||||
Db 1050 GTCCCCCTGGGAAGAGGAGGAGCCAGGCGCTC 1082

RESULT 15
US-09-964-824A-255
; Sequence 255, Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:
; APPLICANT: Horrikan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964,824A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 255
; LENGTH: 2542
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-255

Query Match 7.8%; Score 106; DB 10; Length 2542;
Best Local Similarity 52.1%; Pred. No. 1.le-17;
Matches 236; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

QY 122 TACAAGGTCCACCGGCCCCAGGGTCCCAAGAGGTGACAGAGGATCCCAAGGACCCCTG 181
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 630 TCCCTGGACACCGGGTCCCGAGGCATCAGGGGTACCCAGGCATGCGAGGGGCCCAAGG 689

QY 182 GCCCAACTGGCAACAGGACAGAAAGGAGAGAGGAGGAGCGCTGGACCACTGGCCCTG 241
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 690 GAGAGCGGCCCTCATGGATATAAGGCATGCTGGGCGCTATCGTGCCACTGGGCCAC 749

QY 242 CGGGTGAGAGGCCCAATTGGACAGCTGTTCCCGGAGAGCGTGGCGGCAAGGAT 301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 750 CGGGTGAGGAGGTCTTAGGGACCGCCAGGCGGAGCTGGGAGAAAGGTGACGAGGCA 809

QY 302 CTAAAGGCTCCAGGCGCCCAAGGCTCCGTTGTTCCCTGGGAAGCCCGGCCCTCAGG 361
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 810 GCCCAGGTATTCTGTGACCCCGAGGGGATCAGAGGCCGAAAGGAGCAACGGGCCCCAG 869

QY 362 GCCCCAGTGGGACCCAGGCCCGCCCGGCCACCAAGGAGGAGTCCCGGCCCTC 421
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 870 GCATCAACGGCAAGCATGGACCCCGAGCCCTGGCATGAAGGCGACTGACGACAGG 929

QY 422 AGGGCCCTCTGGCTTCAGGAGCTTCAGGGACCCGTTGGGAGCGCTGGGTGCTTGAC 481
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 930 CGGGACAGCCCGGAAGTCCAGGCCACAGGGGCTAGCGGCTGTCAGGCCAGCCTGGGA 989

QY 482 CTCGGGAGTCCAGGCTTGGCTGGGTACCAGGCATGCCAGGCCCAAGGCCCGCCG 541
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 990 CAAAAGGAGGCCCTTGAGACAGGCTGAGCGGGGCCCGAGGGCCCTTCTGGATTCTG 1049

QY 542 GCCCTCCTGGCCCATCAGGAGCGGTGGTGCCTC 574
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1050 GTCCCCCTGGGAAGAGGAGGAGCCAGGCGCTC 1082
```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2003, 01:34:06 ; Search time 1777.64 Seconds
(without alignments)
12344.923 Million cell updates/sec

Title: US-09-763-712A-1_COPY_670_2024
Perfect score: 1355
Sequence: 1 atgaacaagattgatgag.....agaaaaaggtatcatcccg 1355

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_Other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES | | | |
|------------|-------|--------------|--------------------|
| Result No. | Score | Match Length | Description |
| 1 | 900 | 66.4 | BC009162 Mus muscu |
| 2 | 762.4 | 56.3 | AL568743 |
| 3 | 643.8 | 47.5 | BQ674807 AGENCOURT |
| 4 | 556.4 | 41.1 | BQ713873 AGENCOURT |
| 5 | 523.4 | 38.6 | BQ004176 UI-H-EIO- |
| 6 | 514.6 | 38.0 | AW958053 EST370123 |

| | | | | | | |
|---|----|-------|------|-----|----|----------|
| c | 7 | 499.2 | 36.8 | 655 | 12 | BG482931 |
| | 8 | 469.8 | 34.7 | 500 | 14 | BM676508 |
| | 9 | 467.4 | 34.5 | 507 | 14 | BM713891 |
| | 10 | 462.6 | 34.1 | 906 | 13 | BI456109 |
| | 11 | 445.6 | 32.9 | 580 | 12 | BE910803 |
| | 12 | 441.8 | 32.6 | 808 | 14 | BQ771366 |
| c | 13 | 416.8 | 30.7 | 517 | 9 | AL543000 |
| | 14 | 415.8 | 30.8 | 601 | 10 | BE290299 |
| | 15 | 406 | 30.0 | 638 | 10 | BB248064 |
| | 16 | 403.4 | 29.8 | 957 | 14 | BQ955927 |
| | 17 | 353.4 | 26.1 | 504 | 14 | R74387 |
| c | 18 | 332.4 | 24.5 | 656 | 14 | BQ021066 |
| c | 19 | 332.8 | 23.5 | 345 | 9 | A1742661 |
| | 20 | 318.4 | 23.5 | 352 | 10 | AW956633 |
| | 21 | 318.4 | 23.5 | 552 | 10 | AW956627 |
| | 22 | 314.6 | 23.2 | 421 | 14 | R78202 |
| | 23 | 314.2 | 23.2 | 455 | 10 | AW023880 |
| | 24 | 311 | 23.0 | 339 | 9 | AA304251 |
| | 25 | 308.2 | 22.7 | 943 | 14 | BQ887163 |
| | 26 | 301.4 | 22.2 | 357 | 10 | AW379436 |
| | 27 | 297 | 21.9 | 936 | 14 | BQ891432 |
| c | 28 | 296.4 | 21.9 | 334 | 10 | AW002141 |
| | 29 | 296 | 21.8 | 310 | 9 | AA361740 |
| c | 30 | 289.6 | 21.4 | 318 | 12 | BE767950 |
| | 31 | 285.8 | 21.1 | 309 | 14 | W76214 |
| | 32 | 271 | 20.0 | 350 | 9 | AA012704 |
| | 33 | 268.8 | 19.8 | 294 | 9 | AA357879 |
| c | 34 | 259.6 | 19.2 | 288 | 14 | W72977 |
| | 35 | 254.6 | 18.8 | 414 | 14 | R27402 |
| c | 36 | 252.6 | 18.6 | 498 | 9 | AL569543 |
| c | 37 | 249 | 18.4 | 415 | 14 | R74388 |
| | 38 | 245.8 | 18.1 | 420 | 12 | BE929813 |
| | 39 | 236.2 | 17.4 | 441 | 10 | BB749743 |
| c | 40 | 233.8 | 17.3 | 429 | 12 | BE929798 |
| | 41 | 229.4 | 16.9 | 608 | 10 | AW918393 |
| c | 42 | 222.8 | 16.4 | 381 | 12 | BE767917 |
| c | 43 | 221 | 16.3 | 508 | 9 | AA700864 |
| c | 44 | 220.4 | 16.3 | 538 | 9 | A1954661 |
| c | 45 | 209 | 15.4 | 501 | 10 | AW275816 |

ALIGNMENTS

RESULT 1
BC009162
LOCUS BC009162 3305 bp mRNA linear HTC 07-AUG-2002
DEFINITION Mus musculus, Similar to scavenger receptor with C-type lectin, clone IMAGE:2811487, mRNA.
ACCESSION BC009162
VERSION BC009162.1 GI:14714370
KEYWORDS HTC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3305)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys .
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 5 Row: h Column: 10
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis
 This clone has the following problem: incomplete processing.

| | | |
|------------|--|--|
| FEATURES | Location/Qualifiers | |
| source | 1. .3305 | |
| | /organism="Mus musculus" | |
| | /db_xref="taxon:10090" | |
| | /map="C57BL/6J" | |
| | /clone="IMAGE:2811487" | |
| | /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue." | |
| | /clone_lib="NCI_CGAP_Mam5" | |
| | /lab_host="DH10B" | |
| | /note="Vector: pCMV-SPORT6" | |
| BASE COUNT | 1031 a 757 c 760 g 757 t | |
| ORIGIN | | |

Query Match 66.4%; Score 900; DB 11; Length 3305;
Best Local Similarity 81.7%; Pred. No. 4.5e-173;
Matches 1106; Conservative 0; Mismatches 230; Indels 18; Gaps 5;

| | | | |
|----|------|--|------|
| QY | 1 | ATGCAACAAGATTGTAGAGTTCGAGGTTTACAGACTGAAGTAGCCAACTTATCAGTGATT | 60 |
| Db | 1296 | ATGCAGCAACACATGATGAGGTCAAAGTTTACACTTGAAGTGGCCAACCTTATCAGTGGTT | 1355 |
| QY | 61 | ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGATTTTACAAATA | 120 |
| Db | 1356 | ATGGAAGAGATGAACATTGGTTGACTCCAGGACCGGTGAGCTCATCAAGAACTTTTACCATT | 1415 |
| QY | 121 | CTACAAGGTCACCGGGCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCCAGGGACCCCTT | 180 |
| Db | 1416 | CTACAAGGTCCTCTCGGCCCCAGAGGTCCTCAAAGGTGCACAGAGGATCTCAGGGACCACTT | 1475 |
| QY | 181 | GGCCCAACTGGCAACAGGGACAGAAAGAGAGAGGGGGAGCCTGGACACNCTTGGGCCCT | 240 |
| Db | 1476 | GGTCCAACTGGCAACAAGGACAGAAAGAGAGAGGGAGAGCCTGGTCCACCTTGGCCCT | 1535 |
| QY | 241 | CGGGGTGAGAGAGGCCCAATTGGACCAGCTGTGTCCCGCCGAGAGCGTGGCGGCAAAAGGA | 300 |
| Db | 1536 | CGGGGTGAGAGGGGCACAAATTGNACCAGTGTGGCCCTCTCTGGAGAGCGTGGCAGCAAAAGGA | 1595 |
| QY | 301 | TCTAAAGGCTCCCAAGGGCCCCAAAGGCTCCCGTGGTTTCCCTTGGGAAGCCCGGCCCTCAG | 360 |
| Db | 1596 | TCCAAAGGCTCACAGGCTCCCAAGGATCTCGTGGGTCCCGAGGGAAGCCTTGGCCCTCAA | 1655 |
| QY | 361 | GGCCCCAGTGGGACACAGGCCCCCGGGGCCCAACCGCAAGGCAAGAGGACATCCCGGCCCT | 420 |
| Db | 1656 | GGACCTAGTGGGACCCAGGACCAACAGGTTCACCGAGGAAGGATGGACTCCTTGGGCCCT | 1715 |
| QY | 421 | CAGGGCCCTCTGTGCTTCCAGGACTTCAGGGCACCTTTGGGGAGCCTTGGGGTGCCTTGA | 480 |
| Db | 1716 | CAGGGCCCTCTGTGCTTCCAGGACTACAGGGCACTCTGGGTGAGCCTTGGAGTACCTTGA | 1775 |
| QY | 481 | CTTCGGGGACTTCGAGGCTTGCCTTGGGGTACAGGCAATGCCAGGCCCAAGGGCCCCCCC | 540 |
| Db | 1776 | CTTCGGGGGTTCGAGGCTTTCGCAAGGGGTGCCAGGCATGCTTGGGCCCTAAGGGACCACT | 1835 |
| QY | 541 | GGCCCTCTTGGCCCATCAGAGCGGTGGTGGCCCTTGGCCCTGCAAGATGAGCCAAACCCG | 600 |
| Db | 1836 | GGCCCTCCAGGCCCTCAGAGCAATGAGGCATTGGCTCTGCAAGATGAACCAACCCCA | 1895 |
| QY | 601 | GCACGGGAGGACAATGGCTGCCCGCCCTCACTGGGAAGAACTTCACAGACAAATGCTACTAT | 660 |
| Db | 1896 | GCATCAGAGTCAAGGATGTCACCTTCATCTGGAAGAACTTCACAGATAATGCTACTAT | 1955 |

| | | | |
|----|------|--|------|
| Qy | 661 | TTTTTCAGTTGAGAAAGAAATTTTGTAGGATGCAGAAAGCTTTTCTGTGGAAGACAGATGCTTCCA | 720 |
| Db | 1956 | TTTTTCATTGGAAAAAGAAATTTTGAAGATGCTAAAGCTTTTCTGTGGAAGACAAATCTTCC | 2015 |
| Qy | 721 | CATCTGTTTTCATAACACTAGAGAGGAACAGCAATGATGATAAAAAACAGATGATGGG | 780 |
| Db | 2016 | CATCTCGTTTTCATTAACCTCAGAGGAAGAACAGCAATGATGATAAAAAACATACCGTGGG | 2075 |
| Qy | 781 | AGAGAGAGCCACTGGATCGGCTTCACAGACTCAGAGCGTGAAGATGAATGGAAGTGGCTG | 840 |
| Db | 2076 | AGAGAAAGCCATTGGATCGGCTTCACAGACTCAGAACAGGAAAGCGAATGGAAGTGGCTA | 2135 |
| Qy | 841 | GATGGGCATCTCCAGACTACAAAATTGGAAGCTGGACGCCGATTAAGCTGGGTCAT | 900 |
| Db | 2136 | GACGGTGCACCTGTTGATTACAAAACCTGGAAGCTGGACCAACCAAGATAACTGGGCGAGT | 2195 |
| Qy | 901 | GGCATGGGCGAGGAGAAGACTGTGCTGGGTTGATTTATGCTGGGCAGTGGAAAGATTTTC | 960 |
| Db | 2196 | GGCATGGGCGAGGAGAAGACTGTGCTGGCTTTGATTTACGCAGGACAGCTGGAATGACTTTC | 2255 |
| Qy | 961 | CAATGTGGAAGACGTCAAATAACTTTCATTTCCGAAAAAAGACAGGGAGACAGACTACTGTCATCT | 1020 |
| Db | 2256 | CAGTGTGATGAATCAATAACTTTCATTTCTGTGAGAAGAAAGGAGGCGAGTACCACATPCC | 2315 |
| Qy | 1021 | GCATTATAAGCGACTGTGATGGGATCACATGAGCAAAATTTTCA--GCPTCTCAAAAGGCAAA | 1078 |
| Db | 2316 | ATATTATAGCAGCATGATATAGTACGAAACATATTTTCTGATGCCCTCTGAAAGCCGAA | 2375 |
| Qy | 1079 | GGACACTCCTTTCTAAATTGATCACCCTTCTCATCAGATTGAAAAAAGAAAGCACTGAA | 1138 |
| Db | 2376 | GAATGCTCGTTCTTGATTTCCATCA-CTTCTCACCAGATTG--AATGGAAAAAGCTCTGAA | 2432 |
| Qy | 1139 | AACCAATTACTGAAAAAAATTCAGACGCTAGTGTGTTTTTACCATCCGTCATTACCCCAAG | 1198 |
| Db | 2433 | AGTAGTTATTCAAAATAAATGGACACCTACTG-----CACAAATAACCCAGG | 2480 |
| Qy | 1199 | ACTTGGGAATCAAAATG-TTCCCCAGGGTGATATGCTGATTTTTCATVTGTGCACATGGACT | 1257 |
| Db | 2481 | ACTAGGGGGCTAAATGCTACCCCAAGTTGATATATATGATTTTCCAGTGTAGCAAAATGGACT | 2540 |
| Qy | 1258 | GAATCACATAGATTCTCCTCCGTCAGTACCCGTGGATTATACAAATTAATGTGCTTCCAAA | 1317 |
| Db | 2541 | GAATCGCATGATTTTCTCAGCCATAACCATAGAAATTTATGCAAAATTAATCTTTCCAA | 2600 |
| Qy | 1318 | GATGAACACTCCCAATCAGAAAAAGGTTTATCAT | 1351 |
| Db | 2601 | ATATGGAATCCTCCAATCAGAAAAAGACTATCAT | 2634 |

RESULT 2
 AL568743/c
 LOCUS
 DEFINITION LTI_FL002_PL1 Homo sapiens cDNA clone CS0DE05YH04 3 prime
 mRNA sequence.
 accession AL568743
 version AL568743.1
 keywords GI:12993387
 source EST.
 organism Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 967)
 Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1..967
 /organism="Homo sapiens"
 source

| | | | |
|---|--|--|-------|
| Db | 241 | CTAGAGGACAGCAATGGATATAAAAAACAGATGGTAGGGAGAGAGCCACTGGATCG | 300 |
| QY | 800 | GCCTCACAGACTCAGAGCGTGAATAATGAATGGAAGTGGCTGGATGGGACATCTCCAGACT | 859 |
| Db | 301 | GCCTCACAGACTCAGAGCGTGAATAATGAATGGAAGTGGCTGGATGGGACATCTCCAGACT | 360 |
| QY | 860 | ACAAAAATTGGAAGCTGGACAGCGGATAACTAGGGGTCAATGGCCATGGGCCAGGAGAAG | 919 |
| Db | 361 | ACAAAAATTGGAAGCTGGACAGCGGATAACTAGGGGTCAATGGCCATGGGCCAGGAGAAG | 420 |
| QY | 920 | ACTGTGCTGGGTGATTATATGCTGGGCAGTGGACGATTTCCAAATGTGAAGACGTCAATA | 979 |
| Db | 421 | ACTGTGCTGGGTGATTATATGCTGGGCAGTGGACGATTTCCAAATGTGAAGACGTCAATA | 480 |
| QY | 980 | ACTTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCTATCTGCATATATAACGGACTGTGA | 1039 |
| Db | 481 | ACTTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCTATCTGCATATATAACGGACTGTGA | 540 |
| QY | 1040 | TGGGATCACATGAGCAAAATTTTCAGCTCTCAAGGCAAAAGGACATCTCTTTCTAATTGCA | 1099 |
| Db | 541 | TGGGATCACATGAGCAAAATTTTCAGCTCTCAAGGCAAAAGGACATCTCTTTCTAATTGCA | 600 |
| QY | 1100 | TCACCTTCTCATCAGATTGAAAAAAGCAACACTGAAACCAATTTACTGAAAAAAA | 1158 |
| Db | 601 | TCACCTTCTCATCAGATTG-AAAAAAGCAACACTGAAACCAATTTACTGAAAAAAA | 658 |
| RESULT 4 | | | |
| B0713873 | | | |
| LOCUS | B0713873 861 bp mRNA linear EST 16-JUL-2002 | | |
| DEFINITION | AGENCOURT_8482828 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6306240 | | |
| 5', mRNA sequence. | | | |
| ACCESSION | B0713873 | | |
| VERSION | B0713873.1 GI:21852772 | | |
| KEYWORDS | EST. | | |
| SOURCE | house mouse. | | |
| ORGANISM | Mus musculus | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| TITLE | NIH-MGC http://mgi.nci.nih.gov/ . | | |
| JOURNAL | National Institutes of Health, Mammalian Gene Collection (MGC) | | |
| COMMENT | Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: Susan L. Sullivan, PhD. cDNA Library Preparation: ResGen, Invitrogen Corp cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM13721 row: n column: 01 High quality sequence stop: 598. | | |
| FEATURES | | | |
| source | | | |
| 1..861 | | | |
| /organism="Mus musculus" | | | |
| /db_xref="taxon:10090" | | | |
| /clone="IMAGE:6306240" | | | |
| /clone_lib="NIH_MGC_129" | | | |
| /lab_host="DH10B (phage-resistant)" | | | |
| /note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1.ccdB; Site_1: EcorV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library." | | | |
| BASE COUNT | 221 a | 253 c | 239 g |
| ORIGIN | 147 t | 1 others | |
| Query Match 41.1%; Score 556.4; DB 14; Length 861; | | | |
| Best Local Similarity 82.8%; Pred. NO. 3.7e-103; | | | |
| Matches 647; Conservative 0; Mismatches 132; Indels 2; Gaps 1; | | | |

| | | | |
|------------|--|--|-----|
| QY | 14 | TGATAGGTGTCGAGGTTAGACACTGAAGTAGCCAACTTTATCACTGATTATGGAAGAAATGA | 73 |
| Db | 14 | TGATAGGTCAAAGTTAGACACTGAAGTGGCCAACTTTATCACTGGTTATGGAAGAGATGA | 73 |
| QY | 74 | AGCTAGTAGACTCCAAGCATGCTCAGCTCATCAAGAATTTTACAATACTACAAGGTCCAC | 133 |
| Db | 74 | AACTGGTTGACTCCAAGCAGGTGACGCTCATCAAGAATTTTACCATTCTACAAGGTCCCT | 133 |
| QY | 134 | CGGGCCCCAGGGTCCAAGAGGTGACAGAGGATCCCAGGAGCCCTCGGCCCAACTGGCA | 193 |
| Db | 134 | CTGGCCCCAGAGTCCAAAGGTGACAGAGATCTCAGGGACCCTGTGTCCAATGGCA | 193 |
| QY | 194 | ACAAGGGACAGAAAGAGAGAGAGGGAGCCTGGAGCACCTGGCCCTCGGGGTGAGAGAG | 253 |
| Db | 194 | ACAAGGGACAGAAAGAGAGAGAGGGAGAGCCTGGTCCACCTGGCCCTGCGGGTGAGAGG | 253 |
| QY | 254 | GCCCAATTTGGACCAGCTGGTCCCCCGGAGAGCGTGGCGGCAAGGATCTAAGGCTCCC | 313 |
| Db | 254 | GCACAATTTGGACCATGCGGCCCTCTGGAGAGCGTGGCAGCAAGGATCCAAAGGCTCAC | 313 |
| QY | 314 | AGGGCCCCAAAGGCTCCCGTGGTTCCCTTGGGAAGCCGGGCCCTCAGGGCCCGAGTGGG | 373 |
| Db | 314 | AGGTGCCCAAGAGATCTGTGGGTCCCGAGGAGCGCTGGCCCTCAAGGACCTAGTGGG | 373 |
| QY | 374 | ACCCAGGCCCCCGGGCCACAGGCAAAAGAGGAGTCTCCCGGGCCCTCAGGGCCCTCTCTG | 433 |
| Db | 374 | ACCCAGGACCACAGGTCCACCAGGCAAGGATGGACTCCCTGGCCCTCAGGGCCCTCTG | 433 |
| QY | 434 | GCTTCCAGGGACTTCAGGCGACCGTTGGGGAGCCTTGGGGTGGCTGGACCTCGGGGACTGC | 493 |
| Db | 434 | GCTTCCAGGGACTTCAGGCGACCTGTGGGTGAGCCTGGAGTACCTGGAGCTCGGGGGTTCG | 493 |
| QY | 494 | CAGGCTTGGCTGGGTACCGAGCATGCCAGGCCCCCAAGGGCCCCCGGCCCTCTCTGGCC | 553 |
| Db | 494 | CAGGCTTGGCAGGGTGGCAGGATGCGTGGCCCTTAAGGGACCCTGGCCCTCCAGGCC | 553 |
| QY | 554 | CATCAGGAGCGGTGGTGGCCCTGCGAGATGAGCCAAACCCGGCACCGGAGGACA | 613 |
| Db | 554 | CCTCAGGAGCAATGAGGCGCATTTGGCTCTGCAAAATGAACCAACCCNANCATCAGAGTCA | 613 |
| QY | 614 | ATGGCTGCCCGCTCACCTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGACA | 673 |
| Db | 614 | ACGGATGTCGGCTCACCTGGAAGAACTTCACAGATAAATGCTACTATTTTCTTGGCNA | 673 |
| QY | 674 | AAGAAATTTTGGAGTGCACAAAGC--TTTCTGTGGAAGACAAAGTCTTCACATCTTGT | 731 |
| Db | 674 | AAAAAATTTTGGAAAATGCTACCCCTTTTCTGCGAACAATAATCTCCCATCCGGTTT | 733 |
| QY | 732 | CATAAAGCTAGAGAGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGAGCA | 791 |
| Db | 734 | CATAACCCCAAGAAAAACGCCATGGGATAAAAAACCTACCTTCGGGGACCAAGAAC | 793 |
| QY | 792 | C 792 | |
| Db | 794 | C 794 | |
| RESULT 5 | | | |
| BQ004176/c | | | |
| LOCUS | BQ004176 723 bp mRNA linear EST 26-MAR-2002 | | |
| DEFINITION | UI-H-EIO-aym-p-04-0-UI.sl NCI-CGAP_EIO Homo sapiens cDNA clone | | |
| IMAGE | IMAGE:5840883 3', mRNA sequence. | | |
| ACCESSION | BQ004176 | | |
| VERSION | BQ004176.1 | | |
| KEYWORDS | EST. GI:19729076 | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| TITLE | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index | | |

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES

source
Location/Qualifiers
1..723

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5840883"
/clone_lib="NCL_CGAP_E10"
/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Left Pelvis; Vector: pT7T3-Pac (Pharmacia)
with a modified polylinker; Site:1. EcoR I; Site:2. Not I;
NCL_CGAP_E10 is a cDNA library containing the following
tissues): Chondrosarcoma. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is ACACCTGCAC.
TAG_LIB=UI-H-E10
TAG_TISSUE=chondrosarcoma
TAG_SEQ=ACACTTGCAC"

BASE COUNT 185 a 148 c 130 g 260 t

Query Match 38.6%; Score 523.4; DB 14; Length 723;
Best Local Similarity 99.5%; Pred. No. 1.9e-96;
Matches 546; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
QY 803 TCACAGACTCAGAGCGTGAATAATGAATGGTGGTGGATGGGACATCTCCAGACTACA 862
|||||
Db 723 TCACAGACTCAGAGCGTGAATAATGAATGGTGGTGGAT-GGACATCTCCAGACTACA 665
QY 863 AAAATTGGAAGCTGGACAGCCGGATACTGGGGTCAATGGCCATGGCCAGGAGAGACT 922
|||||
Db 664 AAAATTGGAAGCTGGACAGCCGGATACTGGGGTCAATGGCCATGGCCAGGAGAGACT 605
QY 923 GTGCTGGTGTGATTTATGCTGGCAGTGAACGATTTCCAATGTGAAGACGTCAATAACT 982
|||||
Db 604 GTGCTGGTGTGATTTATGCTGGCAGTGAACGATTTCCAATGTGAAGACGTCAATAACT 545
QY 983 TCATTTCGAAAAAGACAGGAGACAGTACTGTCATCTGCATTATAACGGACTGTGATGG 1042
|||||
Db 544 TCATTTCGAAAAAGACAGGAGACAGTACTGTCATCTGCATTATAACGGACTGTGATGG 485
QY 1043 GATCATATGAGCAAAATTTTCAGCTCTCAAAGGCAAGGACATCCCTTTCTAATTCATCA 1102
|||||
Db 484 GATCATATGAGCAAAATTTTCAGCTCTCAAAGGCAAGGACATCCCTTTCTAATTCATCA 425
QY 1103 CCTTCTCATCATGTAAGAAAAAAGACACTGAAACCAATTAAGAAAAAATTTGA 1162
Db 424 CCTTCTCATCATGTTG-AAAAAAAGACACTGAAAGCCAAATTAAGAAAAAATTTGA 366
QY 1163 CAGCTAGTGTGTTTTTACCATCGTTCATTACCAAGACACTTGGGAACATAAATGTTCCCA 1222
Db 365 CAGCTAGTGTGTTTTTACCATCGTTCATTACCAAGACACTTGGGAACATAAATGTTCCCA 306
QY 1223 GGGTGATATGCTGATTTTCATTTGTGCACATGGACTGAATCACAATGATTTCTCCTCGTCA 1282
|||||

Db

305 GGGTGATATGCTGATTTTCATTTGTCACATGGACTGAATCACAATGATGTTCTCCTCGTCA 246
QY 1283 GTAACCGTGCAGATTATACAAATTTATGCTTTCAAAGTATGGAACACTTCCAATCAGAAAAA 1342
|||||
Db 245 GTAACCGTGCAGATTATACAAATTTATGCTTTCAAAGTATGGAACACTTCCAATCAGAAAAA 186
QY 1343 GCTTATCAT 1351
|||||
Db 185 GCTTATCAT 177
|||||

RESULT 6
AW958053

LOCUS
DEFINITION EST370123 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION AW958053
VERSION AW958053.1 GI:8147736
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 552)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.

TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray

JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 115

Seq primer: Reverse.

Location/Qualifiers

1..552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGE"
/note="Vector: pBluescriptSKm"

BASE COUNT 176 a 112 c 143 g 121 t

FEATURES

source
Location/Qualifiers
1..552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGE"
/note="Vector: pBluescriptSKm"

Query Match 38.0%; Score 514.6; DB 10; Length 552;
Best Local Similarity 98.0%; Pred. No. 1.2e-94;
Matches 542; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

QY 574 CTGGCCCTCAGAAATGAGCCACCCCGGACGAGGACAAATGGCTGCCCTCACTGG 633

Db 1 CTGGCCCTCAGAAATGAGCCACCCCGGACGAGGACAAATGGCTGCCCTCACTGG 60

QY 634 AAGAATCTCACAGACAAATGCTACTATTTTTCAGTTTGAGAAAGAAATTTTGAGGATGCA 693

Db 61 AAGAATCTCACAGACAAATGCTACTATTTTTCAGTTTGAGAAAGAAATTTTGAGGATGCA 120

QY 694 AAGCTTTTCTGTGAGACAAAGTCTTCACATCTTGTGTTTTCATAAACAAGTGTGAGGAGAACAG 753

Db 121 AAGCTTTTCTGTGAGACAAAGTCTTCACATCTTGTGTTTTCATAAACAAGTGTGAGGAGAACAG 180

QY 754 CAATGGATAAAAAACAAGATGGTAGGAG 813

Db 181 CAATGGATAAAAAACAAGATGGTAGGAG 240

QY 814 GAGCGGTGAAGTGAAGTGGCTGGATGGGACATCTCCAGACTTACAAAAATTTGCAAA 873

Db 241 GAGCGGTGAAGTGAAGTGGCTGGATGGGACATCTCCAGACTTACAAAAATTTGCAAA 300

QY 874 GCTGGACAGCCGGATAACTGGGGTCAATGGCCATATGGCCAGGAGAGAGAGAGAGAGAGAGAGAGAG 933

Db 301 GCTGGACAGCCGGATAACTGGGGTCAATGGCCATATGGCCAGGAGAGAGAGAGAGAGAGAGAGAGAG 960

```
QY 934 ATTATGCTGGCAGTGAACGATTTCCAAATGTGAAGACGTCAATAAC-TTCATTTGCGA 992
|||||
Db 361 ATTATGCTGGCAGGGAACGATTTCCAAATGTGAAGACGTCAATAAC-TTAATTGCGA 420
|||||
QY 993 ARAAGACAGGAGACAGTACTGTCATCTGCATATTAACGGACTGTGATGGATCATGA 1052
|||||
Db 421 ARAAGACAGGAGACAGTACTGTCATCTGCATATTAACGGACTGTGATGGATCATGA 480
|||||
QY 1053 GCAAAATTTTCAGCTCTCAAGGCAAGGACACTCCCTTTCTAATTGTCATCACCTTTCATC 1112
|||||
Db 481 GCAAAATTTTCAGCTCTCAAGGCAAGGAC-CTCCCTTCTAATTGGCTCACCTTTCATC 539
|||||
QY 1113 AGATTGAAAAAA 1125
|||||
Db 540 AGATGAAAAAAA 552
|||||

RESULT 7
BG482931
LOCUS 602502939F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4616625 5',
DEFINITION mRNA sequence.
ACCESSION BG482931
VERSION BG482931.1 GI:13415210
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 655)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapb@remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1371 row: m column: 10
High quality sequence stop: 651.
Location/Qualifiers
1. .655
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:4616625"
/clone_lib="NIH_MGC_77"
/note="Organ: lung; Vector: pDNR-LTB (Clontech); Site:1:
sfll (ggccgctcgcc); Site:2: sfll (ggccattatggc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCCATATGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGGCGGCGGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: This is a NIH_MGC Library."
BASE COUNT 219 a 124 c 144 g 167 t 1 others
ORIGIN

Query Match 36.8%; Score 499.2; DB 12; Length 655;
Best Local Similarity 98.1%; Pred. No. 1.6e-91;
Matches 516; Conservative 0; Mismatches 8; Indels 2; Gaps 1;

QY 826 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCG 885
|||||
Db 1 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCG 60
|||||
```

```
QY 886 GATAACTGGGGTCATGGCCATGGCCAGGAGAAAGACTGCTGGGGTTGATTTATGCTGGG 945
|||||
Db 61 GATAACTGGGGTCATGGCCATGGCCAGGAGAAAGACTGCTGGGGTTGATTTATGCTGGG 120
|||||
QY 946 CAGTGAAGCAGTTTCAATGTGAAGACGTCAATAACTTTCATTTGCCAAAAAGACAGGAG 1005
|||||
Db 121 CAGTGAAGCAGTTTCAATGTGAAGACGTCAATAACTTTCATTTGCCAAAAAGACAGGAG 180
|||||
QY 1006 ACAGTACTGTCATCTGCATTAACGAGCTGTGATGGGATCACATGAGCAAAATTTTCAGC 1065
|||||
Db 181 ACAGTACTGTCATCTGCATTAACGAGCTGTGATGGGATCACATGAGCAAAATTTTCAGC 240
|||||
QY 1066 TCTCAAAGCAAGGACACTCCTTTCTAATTGTCATCACCTTCTCATCAGATTGAAAAAA 1125
|||||
Db 241 TCTCAAAGCAAGGACACTCCTTTCTAATTGTCATCACCTTCTCATCAGATTG--AAAAA 298
|||||
QY 1126 AAAAAAGCACTGAAAAACCAATTACTGAAAAAAATTCACAGCTAGTCTTTTACCACATCG 1185
|||||
Db 299 AAAAAAGCACTGAAAAACCAATTACTGAAAAAAATTCACAGCTAGTCTTTTACCACATCG 358
|||||
QY 1186 TCATTACCCAAAGACTTGGGAACATAAAATGTTCCCCAGGGTGATATGCTGATTTTCATTG 1245
|||||
Db 359 TCATTACCCAAAGACTTGGGAACATAAAATGTTCCCCAGGGTGATATGCTGATTTTCATTG 418
|||||
QY 1246 TGCACATGGACTGAATCACATAGATTCTCTCGTCAGTAACCGTCGGATTATACAAAT 1305
|||||
Db 419 TGCACATGGACTGAATCACATAGATTCTCTCGTCAGTAACCGTCGGATTATACAAAT 478
|||||
QY 1306 ATGTCTTCCAAAGTAGTGAACACACTCCAATCAGAAAAAGGTTATCAT 1351
|||||
Db 479 ATGTCTTCCAAAGTAGTGAACACACTCCAATCAGAAAAAGGTTATCAT 524
|||||

RESULT 8
BM676508/c
LOCUS BM676508
DEFINITION UI-E-EJ0-ahq-c-05-0-UI.s2 UI-E-EJ0 Homo sapiens cDNA clone
ACCESSION BM676508
VERSION BM676508.1 GI:18986404
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 500)
UI-E-EJ0-ahq-c-05-0-UI 3', mRNA sequence.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Forward
POLYA=Yes
Location/Qualifiers
1. .500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="UI-E-EJ0-ahq-c-05-0-UI"
/clone_lib="UI-E-EJ0"
/tissue_type="fetal eyes, lens, eye anterior segment,
```


optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-EJ0 is a subcloned cDNA library constructed according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT; optic nerve, CCATTAAAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).
TAG_LIB=UI-E-EJ0
TAG_TISSUE=human eye anterior segment
TAG_SEQ=AAATGCCGCAT"

| | | | | |
|---------------------------|--|---|------|--------|
| BASE COUNT | 109 a | 123 c | 89 g | 179 t |
| ORIGIN | | | | |
| Query Match | 34.7%; Score 469.8; DB 14; Length 500; | | | |
| Best Local Similarity | 99.4%; Pred. No. 1.5e-85; | | | |
| Matches 482; Conservative | 0; Mismatches 2; Indels 1; Gaps 1; | | | |
| QY | 674 | AAAGAAATTTTGGAGTGAAGTCTTCTGTGAAGACAAAGTCTTACATCTTTTTCATCA | 733 | |
| Db | 494 | AGGAATTTTGGAGTGAAGTCTTCTGTGAAGACAAAGTCTTACATCTTTTTCATCA | 435 | |
| QY | 734 | TAACACTAGAGGAGAACAGCAATGGATATAAAACAGATGGTAGGAGAGAGACCAT | 793 | |
| Db | 434 | TAACACTAGAGGAGAACAGCAATGGATATAAAACAGATGGTAGGAGAGAGACCAT | 375 | |
| QY | 794 | GGATGGCTTCACAGACTCAGACGCGTGAATGAATGAAGTGGCTGGATGGACATCTC | 853 | |
| Db | 374 | GGATGGCTTCACAGACTCAGACGCGTGAATGAATGAAGTGGCTGGATGGACATCTC | 315 | |
| QY | 854 | CAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGTCTATGCCATGGCCAG | 913 | |
| Db | 314 | CAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTGGGTCTATGCCATGGCCAG | 255 | |
| QY | 914 | GAGAGACTGTCTGGGTGATTATATGCTGGCAGTGGAAACGATTTCATGTGAAGACG | 973 | |
| Db | 254 | GAGAGACTGTCTGGGTGATTATATGCTGGCAGTGGAAACGATTTCATGTGAAGACG | 195 | |
| QY | 974 | TCAATAACTTTCATTTGCGAAAAGACAGGAGACAGTACTGTCACTGCATTATACGGA | 1033 | |
| Db | 194 | TCAATAACTTTCATTTGCGAAAAGACAGGAGACAGTACTGTCACTGCATTATACGGA | 135 | |
| QY | 1034 | CTGTGATGGATCAGATGACAAATTTTTCAGCTCTCAAGGCAAGGACACCTCTTCTA | 1093 | |
| Db | 134 | CTGTGATGGATCAGATGACAAATTTTTCAGCTCTCAAGGCAAGGACACCTCTTCTA | 75 | |
| QY | 1094 | ATTGCATACCTTCTCATCAGATTGAAAAAAGCACTGAAACCAATCTGAA | 1153 | |
| Db | 74 | ATTGCATACCTTCTCATCAGATTG-AAAAAAGCACTGAAACCAATCTGAA | 16 | |
| QY | 1154 | AAAAA 1158 | | |
| Db | 15 | AAAAA 11 | | |
| RESULT 9 | | | | |
| LOCUS | BM713891 | | | |
| DEFINITION | UI-E-EJ0-ahq-c-05-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone | | | |
| | UI-E-EJ0-ahq-c-05-0-UI 5', mRNA sequence. | | | |
| | | 507 bp | mRNA | linear |
| | | EST 28-FEB-2002 | | |

| | | |
|---------------------------|--|---|
| ACCESSION | BM713891 | GI:19027149 |
| VERSION | BM713891.1 | |
| KEYWORDS | EST. | |
| SOURCE | human. | |
| ORGANISM | Homo sapiens | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| AUTHORS | 1 (bases 1 to 507) | |
| TITLE | Normalization and subtraction: two approaches to facilitate gene discovery | |
| JOURNAL | Genome Res. 6 (9), 791-806 (1996) | |
| MEDLINE | 97044477 | |
| COMMENT | Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msoares@blue.weeg.uiowa.edu Tissue Procurement: Dr. Gregg Hageman cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com). The following repetitive elements were found in this cDNA sequence: 443-498, >POLY_A#Simple_repeat Seq primer: M13 Reverse. | |
| FEATURES | Location/Qualifiers | |
| source | 1..507 | |
| | /organism="Homo sapiens" | |
| | /db_xref="taxon:9606" | |
| | /clone="UI-E-EJ0-ahq-c-05-0-UI" | |
| | /clone_lib="UI-E-EJ0" | |
| | /tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid" | |
| | /dev_stage="fetal and adult" | |
| | /lab_host="DH10B (Life Technologies) (T1 phage resistant)" | |
| | /note="organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-EJ0 is a subcloned cDNA library constructed according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT; optic nerve, CCATTAAAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)." | |
| BASE COUNT | 184 a | 89 c 121 g 112 t |
| ORIGIN | | |
| Query Match | 34.5%; Score 467.4; DB 14; Length 507; | |
| Best Local Similarity | 99.4%; Pred. No. 4.7e-85; | |
| Matches 479; Conservative | 0; Mismatches 2; Indels 1; Gaps 1; | |
| QY | 677 | AAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAAAGTCTTACATCTTTTTCATAA 736 |
| Db | 1 | AAATTTTTCAGGATGCAAGCTTTTCTGTGAAGACAAAGTCTTACATCTTTTTCATAA 60 |
| QY | 737 | ACACTAGAGGAGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGACCATGGA 796 |
| Db | 61 | ACACTAGAGGAGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGACCATGGA 120 |
| QY | 797 | TGGGCTCAGACTCAGACGCTGAAATGAATGAAGTGGCTGGATGGGACATCTCCAG 856 |

|||||
Db 121 TCAGCTCAGACTCAGAGCGTGAATGAATGAAGTGGCTGGATGGACATCTCCAG 180
QY 857 ACTACAAAATTTGGAAGCTGGACAGCCGGAATACCTGGGTCATGCCCATGGCCAGGAG 916
Db 181 ACTACAAAATTTGGAAGCTGGACAGCCGGAATACCTGGGTCATGCCCATGGCCAGGAG 240
QY 917 AAGACTGTGCTGGTGGTGAATTTATGCTGGCAGTGGAACTTCCCAATGTAAGACGTCA 976
Db 241 AAGACTGTGCTGGTGGTGAATTTATGCTGGCAGTGGAACTTCCCAATGTAAGACGTCA 300
QY 977 ATAACCTTATTTGCGAAAAAGACAGGAGACAGTACTGTCTATCTGCAATTAACGACGTG 1036
Db 301 ATAACCTTATTTGCGAAAAAGACAGGAGACAGTACTGTCTATCTGCAATTAACGACGTG 360
QY 1037 TGATGGATCAGATGAGCAAAATTTTCCAGCTCTCAAGGCAAAAGGACACTCTTCTTAAT 1096
Db 361 TGATGGATCAGATGAGCAAAATTTTCCAGCTCTCAAGGCAAAAGGACACTCTTCTTAAT 420
QY 1097 GCATCACCTTCTCATCAGATTTGAAAAAAGACACTGAAACCAATTAAGCAAAAA 1156
Db 421 GCATCACCTTCTCATCAGATTTGAAAAAAGACACTGAAACCAATTAAGCAAAAA 479
QY 1157 AA 1158
Db 480 AA 481

RESULT 10
BI456109
LOCUS
DEFINITION
603172765F1 NCI_CGAP_Mam5 Mus musculus cdna clone IMAGE:5251888 5',
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BI456109 906 bp mRNA linear EST 21-AUG-2001
603172765F1 NCI_CGAP_Mam5 Mus musculus cdna clone IMAGE:5251888 5',
mRNA sequence.
BI456109
BI456109.1 GI:15246765
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 906)
NIH-MGC <http://mgi.nci.nih.gov/>,
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue procurement: Lothar Hennighausen Ph.D., Robin Humphreys
CDNA Library preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLNL1636 row: b column: 17
High quality sequence stop: 730.
Location/Qualifiers

FEATURES
source
1..906
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5251888"
/issue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10b"
/note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 247 a 228 c 256 g 175 t

ORIGIN

Query Match
Best Local Similarity 34.1%; Score 462.6; DB 13; Length 906;
Matches 597; Conservative 0; Mismatches 124; Indels 10; Gaps 5;

QY 1 ATGCAACAAGATTTGATGAGGTCGAGGTAGACACTGAAGTAGCCAACTATACAGTGATT 60
Db 123 ATGCAACAAGATTTGATGAGGTCGAGGTAGACACTGAAGTAGCCAACTATACAGTGATT 182
QY 61 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTTCAGCTCATCAAGAATTTACAATA 120
Db 183 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTTCAGCTCATCAAGAATTTACAATA 242
QY 121 CTACAAGGTCTCCAGGCCCCAGGGTCTCAAGAGGTGACAGAGGATCCAGGACCCCCCT 180
Db 243 CTACAAGGTCTCCAGGCCCCAGGGTCTCAAGAGGTGACAGAGGATCCAGGACCCCCCT 302
QY 181 GCGCCCAACTGGCAACAAGGAGCAGAAAGGAGAGAGAGGGGGAGCCCTGGACACCTGGCCCT 240
Db 303 GGTCCAACCTGGCAACAAGGAGCAGAAAGGAGAGAGGGGGAGCCCTGGACACCTGGCCCT 362
QY 241 GCGGGTGTAGAGAGAGGCCCAATTTGGACAGCTGGTTCCTCCCGGAGAGCGTGGCGCAAGGA 300
Db 363 GCGGGTGTAGAGAGAGGCCCAATTTGGACAGCTGGTTCCTCCCGGAGAGCGTGGCGCAAGGA 422
QY 301 TCTAAAGGCTCCAGGCCCCAGGGTCTCCAGGCTCCGCTGGTTCCTCCCGGAGAGCGTGGCGCTCAG 360
Db 423 TCCAAAGGCTCCAGGCCCCAGGGTCTCCAGGCTCCGCTGGTTCCTCCCGGAGAGCGTGGCGCTCAG 482
QY 361 GCGCCCACTGGGAGAGGCCCAATTTGGACAGCTGGTTCCTCCCGGAGAGCGTGGCGCAAGGA 420
Db 483 GCGCCCACTGGGAGAGGCCCAATTTGGACAGCTGGTTCCTCCCGGAGAGCGTGGCGCAAGGA 542
QY 421 CAGGGCCCTCTCGGCTTCCAGGAGCTTTCAGGAGCGTTCAGGAGCGTGGGGAGCGTGGGGTGCCTGA 480
Db 543 CAGGGCCCTCTCGGCTTCCAGGAGCTTTCAGGAGCGTTCAGGAGCGTGGGGTGCCTGA 602
QY 481 CCTCGGGGAGTGGCAGGCTTCTCGGCTTCCAGGAGCGTTCAGGAGCGTGGGGTGCCTGA 540
Db 603 CCTCGGGGAGTGGCAGGCTTCTCGGCTTCCAGGAGCGTTCAGGAGCGTGGGGTGCCTGA 662
QY 541 GCGCCCTCTCGGCTTCCAGGAGCGTTCAGGAGCGTTCAGGAGCGTGGGGTGCCTGA 600
Db 663 GG-CCTCCAGGCCCCCTCAGGAGC-ATGGAGCCATTGGTCTGCAGAAATGAACCAAGCCCA 720
QY 601 GCACCCGAGGACAATGGC--TGCCCGCCCTCAGTGGAGAACTTCACA--GACAATGCT 655
Db 721 GCATCAGAGGTTTCAACGCGATGTCTCTCAGTGGAGAACTTCACGCGATTAAATGGCT 780
QY 656 ACTATTTTTCAGTTG--AGAAAGAAATTTTTCAGGAGTGCAGAACTTCCTGTGTGAAGACA 712
Db 781 ACTATTTTTCAGTTG--AGAAAGAAATTTTTCAGGAGTGCAGAACTTCCTGTGTGAAGACA 840
QY 713 AGTCTTTCACAT 723
Db 841 AATTTTCCAAT 851

RESULT 11
BE910803
LOCUS
DEFINITION
601661855F1 NCI_CGAP_Mam1 Mus musculus cdna clone IMAGE:3962292 5',
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

BE910803 580 bp mRNA linear EST 29-SEP-2000
601661855F1 NCI_CGAP_Mam1 Mus musculus cdna clone IMAGE:3962292 5',
mRNA sequence.
BE910803
BE910803.1 GI:10407765
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 580)
NIH-MGC <http://mgi.nci.nih.gov/>,
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

| | | | | | | | | | | |
|----------|---|-----|--|-----|--|--|--|--|--|--|
| COMMENT | Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1AM9128 row: m column: 13 High quality sequence stop: 580. | | | | | | | | | |
| | Location/Qualifiers 1. .580 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /clone="IMAGE:3962292" /clone_lib="NCI_CGAP_Maml" /tissue_type="tumor, biopsy sample" /dev_stage="10 months, virgin" /lab_host="DH10B" /note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH" | | | | | | | | | |
| | BASE COUNT 159 a 153 c 161 g 107 t | | | | | | | | | |
| | ORIGIN | | | | | | | | | |
| | Query Match 32.9%; Score 445.6; DB 12; Length 580; Best Local Similarity 85.5%; Pred. No. 1.3e-80; Matches 496; Conservative 0; Mismatches 84; Indels 0; Gaps 0; | | | | | | | | | |
| | QY | 290 | CGCGCAAGAGATCTAAAGGCTCCAGGGCCCAAGGCTCCGCTGGTTCCTCGGGAAGC | 349 | | | | | | |
| | Db | 1 | GCAGCAAGAGTCCAAAGGCTCAGAGGTCCCAAGGATCTCGTGGTCCCGAGGAAGC | 60 | | | | | | |
| | QY | 350 | CGGGCCCTCAGGGCCCCAGTGGGGACCCAGGCCCCCGGGCCCCACAGGCAAGAGGGAC | 409 | | | | | | |
| | Db | 61 | CTGGCCCTCAAGGACCTAGTGGGGACCCAGGACCACAGGTCCACAGGCAAGGATGGAC | 120 | | | | | | |
| | QY | 410 | TCCCGGGCCCTCAGGGCCCTCTGTGCTCCAGGGACTTCAGGGACCCCTTGGGGAGCCTG | 469 | | | | | | |
| FEATURES | Db | 121 | TCCCTGGCCCTCAGGGCCCTCTGTGCTCCAGGGACTACAGGGCACTGTGGGTGAGCCCTG | 180 | | | | | | |
| | QY | 470 | GGGTGCTTGACCTCGGGGACTGCCAGGCTTGCTGGGGTACCAGGCATGCCAGGCCCA | 529 | | | | | | |
| | Db | 181 | GAGTACCTGGACCTCGGGGTGCCAGGCTGCCAGGGTGCCAGGCATGCTGGGCCCTA | 240 | | | | | | |
| | QY | 530 | AGGGCCCCCGGCCCTCTGTGGCCCATCAGGAGCGTGTGGCCCTGGCCCTGCAGAAATG | 589 | | | | | | |
| | Db | 241 | AGGGACCACCTGGCCCTCCAGGCCCTCAGGAGCAATGGAGCAATGGCTCTGCAGAAATG | 300 | | | | | | |
| | QY | 590 | AGCCAAACCCGGCAGCAGGACAATGGCTGCCGCCCTCACTGGAGAAGCTTCACAGACA | 649 | | | | | | |
| | Db | 301 | AACCACCCAGCATCAGAGGTCAACGGATGTCGCCCTCACTGGGAAGAACTTCACAGATA | 360 | | | | | | |
| | QY | 650 | AATGCTACTATTTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGAGCTTTTCTGTGAAG | 709 | | | | | | |
| | Db | 361 | AATGCTACTATTTTTCATTGGAAGAAAGAAATTTTGAAGATGCTAAGCTTTTCTGTGAAG | 420 | | | | | | |
| | QY | 710 | ACAAGTCTTCACATCTGTTTTCATATAACACTAGAGAGGAAACAGCAATGGATAAAAAAC | 769 | | | | | | |
| RESULTS | Db | 421 | ACAAATCTTCCCATCTCGTTTTTCATATAACTCAAGAGAAGAACAGCAATGGATAAAAAAGC | 480 | | | | | | |
| | QY | 770 | AGATGTAGGAGAGAGACCCATGATCGGCTTCACAGACTCAGAGCTGGAATGAT | 829 | | | | | | |
| | Db | 481 | ATACCGTGGGAGAGAAAGCCATTGGATCGGCTTCACAGACTCAGAGCAAGGAAAGCGAAT | 540 | | | | | | |
| | QY | 830 | GGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTG | 869 | | | | | | |
| | Db | 541 | GGAAGTGGCTAGACGGGTACCTGTGTGATTACAAAACTG | 580 | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |

| | | | | | |
|--|---|--|------------|--------|-----------------|
| RESULT 12 | BQ771366 | 808 bp | mRNA | linear | EST 26-JUL-2002 |
| LOCUS | UI-M-F10-byu-g-09-0-UI.r1.NIH_BMAP_F10 | Mus musculus | cDNA clone | | |
| DEFINITION | IMAGE:5702432 5', mRNA sequence. | | | | |
| ACCESSION | BQ771366 | | | | |
| VERSION | BQ771366 | | | | |
| KEYWORDS | EST. | | | | |
| SOURCE | BQ771366.1 | GI:21979842 | | | |
| ORGANISM | Mus musculus | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 808) | | | | |
| AUTHORS | NIH-MGC http://mgc.nci.nih.gov/. | | | | |
| TITLE | National Institutes of Health, Mammalian Gene Collection (MGC) | | | | |
| JOURNAL | Unpublished (1999) | | | | |
| COMMENT | Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Dr. Jim Lin, University of Iowa cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov This clone was contributed by the Brain Molecular Anatomy Project (BMAP) | | | | |
| FEATURES | Seq primer: pYX-5. Location/Qualifiers 1. .808 /organism="Mus musculus" /strain="C57BL/6" /db_xref="taxon:10090" /clone="IMAGE:5702432" /clone_lib="NIH_BMAP_F10" /tissue_type="whole brain" /dev_stage="embryo 12.5dpc" /lab_host="DH10B (T1 phage resistant)" /note="Organ: Brain; Vector: pYX- Asc; Site.1: EcoR I; Site.2: Not I. The library was constructed according Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCCACGAC. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator." | | | | |
| BASE COUNT | 270 a | 158 c | 173 g | 206 t | 1 others |
| ORIGIN | | | | | |
| Query Match 32.6%; Score 441.8; DB 14; Length 808; Best Local Similarity 77.5%; Pred. No. 7.6e-80; Matches 604; Conservative 0; Mismatches 157; Indels 18; Gaps 5; | | | | | |
| QY | 576 | GGCCCTGCAGAAATGAGCCAAACCCCGGACCGGAGGACAATGGCTGCCGCCCTCACTGGAA | 635 | | |
| Db | 1 | GGCTCTGCAGAAATGAACCAACCCACGATCAGAGGTCAACGGATGTCGCCCTCACTGGAA | 60 | | |
| QY | 636 | GAACCTTCACAGACAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAA | 695 | | |
| Db | 61 | GAACCTTCAGACAATGCTACTATTTTTCATTGGAAAAAGAAATTTTTCAGAGATGCTAA | 120 | | |
| QY | 696 | GCTTTTCTGTGAAGACAATGCTTTCACATCTCTGTTTTCATATAACACTAGAGAGGAACACA | 755 | | |
| Db | 121 | GCTTTTCTGTGAAGACAATCTTCCATCTCGTTTTTCATATAACTCAAGAGAGAACACGA | 180 | | |
| QY | 756 | ATGGATAAAAAACAGATGGTAGGGAGAGAGAGCCACTGGATCGGCTCACAGACTCAGA | 815 | | |

```
|||||
181 ATGGATAAAAAACATACCGTGGGAGAGAAAGCCATTGGATCGGCCTCACAGACTCAGA 240
QY 816 GCGTGAATATGAATGGAAGTGGTGGAGACATCTCCAGACTACAAAAATTGGAAGC 875
Db 241 ACAGGAAACGAATGGAAGTGGTGGAGACATCTCCAGACTACAAAAATTGGAAGC 300
QY 876 TGGACAGCGCGGTAATCTGGGGTTCATGCGCCATGGCCAGGAGAGACTGTGCTGGGTGAT 935
Db 301 TGGACACACAGATACTGGGGCAGTGGCCATGGCCAGGAGAGACTGTGCTGGGTGAT 360
QY 936 TTATGCTGGGAGTGAACGATTTCAATGTGAAGACGTCAATAACTTCATTTGGCAAAA 995
Db 361 TTACGAGGACAGTGAATGACTTCCAGTGTGATGAATCAATAACTTCATTTGTGAGAA 420
QY 996 AGACAGGGAGACAGTACTGCTCATCTGCATTATTAACGGACTGTGATGGGATCATGAGCA 1055
Db 421 GGAAGGGAGGAGTACCATCATCATTAATATACAGCATGATATTAATAGAGAAACATA 480
QY 1056 AATTTTCA--GCTCTCAAAAGGCAAGACACTCCTTTCTTAATTGATCATCACTTCTCATCA 1113
Db 481 TTTTCTGATGCTCTGAAAGCGGAAGTCTCGTTTGTGATTCATCA-CITCTCACCA 539
QY 1114 GATTGAAAAAAGAACGACATGAAACCAATTAATCTGAAAAAATTAATGACAGCTAGTGT 1173
Db 540 GATTG--AATGAAAAAGCTCTGAAAAGTAGTATTCAAAATAAATGGAACACTACTG-- 595
QY 1174 TTTTACCATCGTCAATTCACCAAGACTTGGGAAGCTTAAATG-TTCCGCCAGGATGATAG 1232
Db 596 -----CACATAACCCAGGACTAGGGGCTTAAATGCTCCCCCAAGTTGATATA 645
QY 1233 CTGATTTTCATTTGTCACATGAGTGAATCAGATAGTATCTCTCGTCAGTAACCGTGC 1292
Db 646 TTGATTTCCAGTCAATGAGTGAATCGCATAGATTTTCTCAGCCATTACCATACA 705
QY 1293 GATTATACAAATATGCTTTCACAAAGTATGGAACACTCAATCAGAAAAAGTTATCAT 1351
Db 706 ATTTATGCAAGATATATCTTTCCAAATATGGAATGCTCCCAATCAGAAAAAGACTATCAT 764

RESULT 13
AL543000/c 517 bp mRNA linear EST 16-FEB-2001
LOCUS AL543000 LTI_FL002_PL1 Homo sapiens cDNA clone CS0DE013YJ08 3 prime
DEFINITION mRNA sequence.
ACCESSION AL543000
VERSION AL543000.1 GI:12875478
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 517)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DE013YJ08"
/lab_host="LTI_FL002_PL1"
/note="Organ: Placenta; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
```

```
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 70 a 117 c 117 g 201 t 12 others
ORIGIN
Query Match 30.8%; Score 416.8; DB 9; Length 517;
Best Local Similarity 89.5%; Pred. No. 9.3e-75;
Matches 451; Conservative 7; Mismatches 45; Indels 1; Gaps 1;
QY 729 TTTTATAAACAACACTAGAGAGGAACACCAATGGATAAAAAACAGATGGTAGGAGAGAGAG 788
Db 517 TTTTATAAACAACACTAGAGAGGAACACCAATGGATAAAAAACAGATGGTAGGAGAGAGAG 458
QY 789 CCACCTGGATCGGCCCTCACAGACTCAGAGCGTCAATGAATGAATGAATGGCTGGTGGGAC 848
Db 457 CCACCTGGATCGGCCCTCACARACTCAGAGCGTGAATGAATGAATGGCTGGTGGGAC 398
QY 849 ATCTCCAGACTACAAAAATTGAAAGCTGGACAGCCGGATAACTGGGTCATGGCCATGG 908
Db 397 ATCTCCAGACTACAAAAATTGAAAGCTGGACAGCCGGATWACTGGGTCATGGCCATGG 338
QY 909 GCACGAGAAGACTGTGCTGGGTTGATTTATGCTGGGAGTGAACGATTTTCCATGTGA 968
Db 337 GCACGAGAAGACTGTGCTGGGTTGATTTATGCTGGGAGTGAACGATTTTCCATGTGA 278
QY 969 AGACGCTCAATAACTTCAATTTGCGAAAAAGACAGGAGACAGTACTGTCTATCTGCATTATA 1028
Db 277 AGACGCTCAATAACTTCAATTTGCGAAAAAGACAGGAGACAGTACTGTCTATCTGCATTATA 218
QY 1029 ACGGACTGTGATGGGATCACATGAGCAAAATTTTCAGCTCTCAAAGCAAAAGGACACTCT 1088
Db 217 ACGGACTGTGATGGGATCACATGAGCAAAATTTAACAGCMCAAAAGGCAAAAGGACACCC 158
QY 1089 TTCTTAATGTCATCACCCTTCTCATCAGATTGAAAAAAGGACACTGAAAAACCAATTAC 1148
Db 157 TCCTAAAAAGCAACACCCCAACCAACAAAG-AAAAAAGGACACCGGCAAAACCAATCAC 99
QY 1149 TGAAAAAATTTGACAGCTAGTGTGTTTTCACCTACCTACCAAAAGACTTGGGAAC 1208
Db 98 CGAAAAAAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAGGACAG 39
QY 1209 TAAATGTGTTCCCGAGGATGATG 1232
Db 38 CAAAGCGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 15

RESULT 14
BE290299 601 bp mRNA linear EST 26-OCT-2000
LOCUS BE290299
DEFINITION BE290299.1 NC1_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3484163 5',
mRNA sequence.
ACCESSION BE290299
VERSION BE290299.1 GI:9171250
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 601)
NIH-MGC http://mgc.nhl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
```


cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCTCGAGTAAATTAATCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pbluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 212 a 134 c 146 g 146 t
ORIGIN

Query Match 30.0%; Score 406; DB 10; Length 638;
Best Local Similarity 80.4%; Pred. No. 1.5e-72;
Matches 513; Conservative 0; Mismatches 120; Indels 5; Gaps 3;

QY 536 CCCCCGGCCCTCTGGCCCATCAGGAGGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCAA 595
DB 1 CACCTGGCCCTCCAGGCCCTCAGGAGCAATGGAGCCATTGGCTCTGCAGAAATGAGCCAA 60
QY 596 CCCCCGACCGGAGGACAAATGCTGCCGCCCTCCTCTGGAAGAACTTCACAGACAAATGCT 655
DB 61 CCCCCAGCATCAGAGGTCAACGGATGCGCCCTCCTCTGGAAGAACTTCACAGATAAATGCT 120
QY 656 ACTATTTTTCAGTTGAGAAAGAAATTTTGAGGATGCAAGCTTTTCTGTGAAGACAAGT 715
DB 121 ACTATTTTTCATGGAAAAAGAAATTTTGAAGATGCTAAGCTTTTCTGTGAAGACAAT 180
QY 716 CTTACATCTTTTTCATAACACTAGAGAGGACAGCAATGATAAAAACAGATGG 775
DB 181 CTTCCATCTCGTTTTCATAAATCAAGAGAGACAGCAATGGATAAAAAGCATACCG 240
QY 776 TAGGAGAGAGAGCCACATGGATCGGCCCTCACAGACTCAGAGCGTGAAATGAATGGAAGT 835
DB 241 TGGGGAGAGAAAGCCATTGGATCGGCCCTCACAGACTCAGACAGGAAAGCGAATGGAAGT 300
QY 836 GCCTGGATGGGACATCTCCAGACTACAAAAATTGAAAGCTGGACAGCCGGATAACTGGG 895
DB 301 GGCTAGACGGGTACCTGTGTGATTACAAAAGCTGGAAGCTGGACACACAGATAACTGGG 360
QY 896 GTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTGATTTATGCTGGCAGTGGAAAG 955
DB 361 GCAGTGGCCATGGCCAGGAGAGACTGTGCTGGCTGATTTACGAGGACAGTGGAAAG 420
QY 956 ATTTTCAATGTGAAGACGTCAATACTTCATTTGGGAAAGACAGAGGAGAGACTACTGT 1015
DB 421 ACTTCCAGTGTGATGAATCAATACTTCATTTGTGAGAAAGGAGGAGGAGGAGTACCAT 480
QY 1016 CATCTGCATTATAAGGACTGTGATGGGATCACATGACAAATTTTCA - GCTCTCAAG 1073
DB 481 CATCCATATTATACAGCATGATATAATAGCAGAAACATATTTTCTGATGCCCTCTGAAG 540
QY 1074 GCAAGGACACTCTTTCTTAATTGATCACCCTTCTCATCAGATTGAAAAAAGCA 1133
DB 541 CGGAAGAAATGCTCGTTTTTGTATTCATCA - CTTCTCACCAGATTG - AATGGAAGAGCT 597
QY 1134 CTGAAACCAATTTACTGAAAAAATTTACAGCTAGT 1171
DB 598 CTGAAAGTAGTTATTCAAAATAAATGGACACCTACTG 635

Search completed: March 21, 2003, 08:08:35
Job time : 1791.64 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 02:20:06 ; Search time 286.911 Seconds
(without alignments)
4293.470 Million cell updates/sec

Title: US-09-763-712A-2

Perfect score: 2969

Sequence: 1 MYSHNVIMNLNLTQVQ.....EDVNNFICKRETVLSSAL 547

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O/Cgn2_1/USPTO_spool/US09763712/runat_14032003_100949_18091/app_query.fasta_1.1877
-DB=N_Geneseq_101002 -QFMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -DOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712.qcgn_1_1_626_0runat_14032003_100949_18091 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Length | DB | ID | Description |
|------------|--------|---------|--------|----|-----------|-----------------------|
| 1 | 2969 | 100.0 | 2024 | 21 | AAA07697 | Human collectin en |
| 2 | 2969 | 100.0 | 2262 | 24 | ABA97932 | Human scavenger re |
| 3 | 2969 | 100.0 | 2628 | 22 | AAH43036 | Nucleotide sequenc |
| 4 | 2963 | 99.8 | 2181 | 22 | AAI60628 | Human polynucleoti |
| 5 | 2963 | 99.8 | 2641 | 22 | AAC66903 | Human EXMAD-14 cod |
| 6 | 2956 | 99.6 | 2318 | 22 | AAI58842 | Human polynucleoti |
| 7 | 2936 | 98.9 | 2929 | 24 | ABQ92072 | Human polynucleoti |
| 8 | 2936 | 98.9 | 2930 | 20 | AAV55746 | Human secreted pro |
| 9 | 2765 | 93.1 | 2637 | 22 | AAH43037 | Nucleotide sequenc |
| 10 | 2688 | 90.5 | 2005 | 24 | ABL95574 | Human angiogenesis |
| 11 | 2688 | 90.5 | 2005 | 24 | ABL88085 | Human PRO7223 CDNA |
| 12 | 2187 | 73.7 | 2256 | 22 | AAH43054 | Nucleotide sequenc |
| 13 | 1715.5 | 57.8 | 1521 | 23 | AA571133 | DNA encoding novel |
| 14 | 765 | 25.8 | 3685 | 20 | AA527858 | Human CSR3 protein |
| 15 | 765 | 25.8 | 3810 | 20 | AA527856 | Human CSR1 protein |
| 16 | 474.5 | 16.0 | 1410 | 24 | ABK84622 | Human CDNA differe |
| 17 | 474.5 | 16.0 | 1410 | 24 | ABL66964 | Lung cancer relate |
| 18 | 471 | 15.9 | 5676 | 15 | AAQ64556 | Human collagen (Ty |
| 19 | 471 | 15.9 | 8284 | 22 | AAAL26526 | Human breast cance |
| 20 | 471 | 15.9 | 8284 | 22 | AAAL26553 | Human breast cance |
| 21 | 471 | 15.9 | 8284 | 22 | AAAL26600 | Human breast cance |
| 22 | 462.5 | 15.6 | 4428 | 22 | AAD06574 | Bovine alphasal(III) |
| 23 | 462.5 | 15.6 | 4428 | 22 | AAD06575 | Bovine alphasal(III) |
| 24 | 459.5 | 15.5 | 756 | 14 | AAQ43034 | Collagen-like poly |
| 25 | 459.5 | 15.5 | 756 | 17 | AAT16768 | Collagen-like poly |
| 26 | 458.5 | 15.4 | 4821 | 23 | AA586866 | DNA encoding novel |
| 27 | 454 | 15.3 | 3171 | 21 | AAAL12503 | CDNA encoding a hu |
| 28 | 454 | 15.3 | 3174 | 21 | AA299843 | DNA encoding human |
| 29 | 453 | 15.3 | 6158 | 24 | ABL62095 | Colon adenocarcino |
| 30 | 453 | 15.3 | 6158 | 24 | ABL65452 | Lung cancer relate |
| 31 | 453 | 15.3 | 6158 | 24 | ABK5486 | Human endometrial |
| 32 | 452.5 | 15.2 | 3170 | 21 | AAAL12493 | CDNA encoding huma |
| 33 | 452.5 | 15.2 | 3171 | 21 | AAAL12502 | CDNA encoding a hu |
| 34 | 452.5 | 15.2 | 3181 | 19 | AAV59358 | Nucleotide sequenc |
| 35 | 452.5 | 15.2 | 3349 | 17 | AAT16518 | Collagen A1/decori |
| 36 | 452.5 | 15.2 | 3349 | 21 | AAAL12500 | CDNA encoding a ch |
| 37 | 452.5 | 15.2 | 3531 | 21 | AAAL12497 | CDNA encoding a ch |
| 38 | 452.5 | 15.2 | 3535 | 17 | AAT16515 | Collagen A1/BMP-2B |
| 39 | 452.5 | 15.2 | 3541 | 17 | AAAL16516 | Collagen A1/TGF-be |
| 40 | 452.5 | 15.2 | 3541 | 21 | AAAL12498 | CDNA encoding a ch |
| 41 | 452.5 | 15.2 | 4192 | 17 | AAT16517 | Collagen A1/decori |
| 42 | 452.5 | 15.2 | 4409 | 19 | AAV60814 | Human recombinant |
| 43 | 452.5 | 15.2 | 4428 | 22 | AAD06578 | Porcine alphasal(III) |
| 44 | 452.5 | 15.2 | 4770 | 22 | AA522441 | Human CDNA encodin |
| 45 | 452.5 | 15.2 | 6728 | 22 | AAF90491 | Human pro-alpha-1 |

ALIGNMENTS

RESULT 1
AAA07697
ID AAA07697 standard; DNA; 2024 BP.
XX AAA07697;
AC
XX
XX 20-JUN-2000 (first entry)
DT
XX Human collectin encoding DNA.
DE
XX Collectin; human; antibacterial; antiviral; ds.
KW
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH CDS 55..1698
FT
FT /*tag= a
FT /product= "collectin"

```
FT misc_feature 1..738 /*tag= b
FT /*note= "fragment specifically claimed in claim 8"
FT misc_feature 55..738 /*tag= c
FT /*note= "fragment specifically claimed in claim 8"
FT misc_feature 79..738 /*tag= d
FT /*note= "fragment specifically claimed in claim 8"
FT misc_feature 325..738 /*tag= e
FT /*note= "fragment specifically claimed in claim 8"
FT misc_feature 358..738 /*tag= f
FT /*note= "fragment specifically claimed in claim 8"
FT misc_feature 670..1695 /*tag= g
FT /*note= "fragment specifically claimed in claim 5"
FT misc_feature 685..738 /*tag= h
FT /*note= "fragment specifically claimed in claim 7"
FT misc_feature 730..738 /*tag= i
FT /*note= "fragment specifically claimed in claim 7"
FT misc_feature 739..1695 /*tag= j
FT /*note= "fragment specifically claimed in claim 6"
FT misc_feature 1696..2024 /*tag= k
FT /*note= "fragment specifically claimed in claim 9"
XX
XX WO200011161-A1.
XX
XX 02-MAR-2000.
XX
XX 24-AUG-1999; 99WO-JP04552.
XX
XX 24-AUG-1998; 98JP-0237611.
XX
XX (FUSO ) FUSO PHARM IND LTD.
XX
XX Wakamiya N;
XX
XX WPI: 2000-224696/19.
XX P-PSDB; AAY77985.
XX
XX New collectin of human origin having antibacterial and antiviral
XX activity, and gene encoding it useful for production of transgenic
XX animals and of antibodies for screening potential drug molecules -
XX
XX Claim 5-9; Page 87-90; 106pp; Japanese.
XX
XX The invention relates to polynucleotides encoding a new collectin of
XX human origin. The collectin can be used as an antibacterial and antiviral
XX agent and for screening potential drug molecules. The new collectin can
XX be produced by standard recombinant methodology. The present sequence
XX represents a DNA encoding the human collectin.
XX
XX Sequence 2024 BP; 595 A; 524 C; 506 G; 399 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 5,17e-96 Length: 2024
XX Score: 2969.00 Matches: 547
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-763-712A-2 (1-547) x AAA07697 (1-2024)
XX
XX 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
XX
XX 55 ATGTATTCTCATATGCTGATCATGACCTCAACACCTGACCTGACCCAGGTGCGAG 114
```

```
Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
Db 115 CAGAGGAACCTCATCCGAACTCGAGCGTCTGTGGATGACACAGCCAGGCTATCCAG 174
Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
Db 175 CGAATCAAGAACGACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGACAGC 234
Qy 61 AspTyrLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80
Db 235 GATTGGCTGAAGGAGAAAGTGCAGAGCTTGCAGAGCGTGGCTGCCAACAACTCTCGGTG 294
Qy 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
Db 295 GCCAAGCCACACAGCACACCTCGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGT 354
Qy 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
Db 355 CAGATGGAGAACATCACCACTATCTCTCAAGCCAAAGCAGCAGACCTGAAAGACCTGCAG 414
Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
Db 415 GACTTACACAAGATGCAGAGAAATAGAACGCCATCAAGTTCAACCACTGGAGGAAGCC 474
Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
Db 475 TTCCAGCTCTTTGAGCGGATATTGTGAACATCATTAGCAATATCATGTACACAGCCAC 534
Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
Db 535 CACCTGCGGAGCGCTGACCAGCAATCTAAATGAAGTCAGGACCACCTTGACAGATACCCTT 594
Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
Db 595 ACCAAACACACAGATGATCTGACCTCCTTGAAATAATACCTTGCCCAACATCCGTTTGGAT 654
Qy 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
Db 655 TCTGTTTCTCTCAGGATGCAACAGATTTGATGAGTGCAGGTTAGACACTCAAGTAGCC 714
Qy 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
Db 715 AACTTATCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATC 774
Qy 241 LysAsnPheThrIleLeuGlnGlyProGlyProGlyProArgGlyProArgLysArgGly 260
Db 775 AAGATTTTACAATACTACAGGTCCACCGGCCCCCAGGGGTCCCAAGAGGTGACAGAGA 834
Qy 261 SerGlnGlyProProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro 280
Db 835 TCCCAAGGACCCCTCGGCCAACCTGGCAACAAGGGACAGAAAGGAGAGAAAGGGAGCCT 894
Qy 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGlu 300
Db 895 GGACCACTGGCCCTGCGGGTGAGAGAGGCCCAATTTGACCAAGCTGGTCCCGCCCGGAGAG 954
Qy 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
Db 955 CGTGGCGCAAGGATCTAAAGGCTCCCGAGGGCCCCCAAGGCTCCCGTGGTCCCTGGG 1014
Qy 321 LysProGlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGlu 340
Db 1015 AAGCCCGGCCCTCAGGGCCCCAGTGGGACCCAGGCCCGCCCGGGGCCACAGGCAAGAG 1074
Qy 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
Db 1075 GGACTCCCGGCCCTCAGGGCCCCCTCTGGCTTCCAGGGACTTCAGGGCACCCTGGGGAG 1134
Qy 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
Db 1135 CCTGGGTGCTTGGACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAGGATGCCAGGC 1194
```


QY 381 ProLysGlyProProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
 Db 1195 CCCAAGGCCCCCGCCCTCTGGCCCATCAGAGCGGTGGTCCCTGCCCTGCGAG 1254
 QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
 Db 1255 AATGAGCCACCCCGCCGACGAGACAAATGGCTGCCCGCTCACTGGAAGAATTCACA 1314
 QY 421 AspLysCysTrpTyrrPheSerValGluLysGluLePheGluAspAlaLysLeuPheCys 440
 Db 1315 GACAAATGCTACTATTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTCTGT 1374
 QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
 Db 1375 GAAGACAAGTCTTCACATCTTGTTCATAAACACACTAGAGAGAAACAAATGGATAAAA 1434
 QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
 Db 1435 AAACAGATGGTAGGAGAGAGAGCCACTGGATCGCCCTCAGACTCAGAGCGTGAAAT 1494
 QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrrLysAsnTrpLysAlaGlyGlnPro 500
 Db 1495 GAATGGAAGTGGCTGCATGGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCG 1554
 QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrrAlaGly 520
 Db 1555 GATAACTGGGTCAATGGCCATGGCCAGGAGAAAGACTGTGCTGGTTGATTATGCTGGG 1614
 QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
 Db 1615 CAGTGGACAGATTTCCAAATGGAAGACGTCATAACTTTTGGCAAAAAGACAGGGAG 1674
 QY 541 ThrValLeuSerSerAlaLeu 547
 Db 1675 ACAGTACTGTCTATCGCATTA 1695
 RESULT 2
 ID ABA97932 standard; cDNA; 2262 BP.
 AC ABA97932;
 XX 25-APR-2002 (first entry)
 DT Human scavenger receptor-like protein encoding cDNA SEQ ID NO 1.
 DE Human; scavenger; receptor; diagnosis; treatment; autoimmune disease;
 KW rheumatoid arthritis; gene; ss.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FH 33..2262
 CDS /*tag= a
 FT /product= "scavenger receptor-like protein"
 XX JP2001340089-A.
 PN 11-DEC-2001.
 PD 08-DEC-2000; 2000JP-0375066.
 PF 27-MAR-2000; 2000JP-0090772.
 PR (SHIO) SHIONOGI & CO LTD.
 PA WPI; 2002-144965/19.
 DR P-PSDB; ABB08642.
 XX New scavenger receptor-like protein for diagnosis, prevention and
 PT treatment of autoimmune disease, such as rheumatoid arthritis -
 XX Claim 3; Fig 1; 38pp; Japanese.

XX The invention relates to a human scavenger receptor-like protein. The
 CC protein is useful as a target molecule for diagnosis, prevention and
 CC treatment of autoimmune diseases such as rheumatoid arthritis.
 SQ Sequence 2262 BP; 670 A; 576 C; 581 G; 435 T; 0 other;
 Alignment Scores:
 Pred. No.: 5.63e-96 Length: 2262
 Score: 2969.00 Matches: 547
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 24 Gaps:
 US-09-763-712A-2 (1-547) x ABA97932 (1-2262)
 QY 1 MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
 Db 619 ATGTATTCTCATAAATGTGGTCAATCATGAACCTCAACAACCTGAACCTGACCAGGTGCAG 678
 QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
 Db 679 CAGAGGAACCTCATCCGAATCTGCAGCGGTCTGTGGATGACCAAGCCAGGCTATCCAG 738
 QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
 Db 739 CGAATCAAGAAGCAGCTTTCAAAATCTGCAGCAGGTTTTTCTCAAGCCAAAGAGACAGC 798
 QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80
 Db 799 GATTGGCTGAAGGAGAAAGTGCAGAGCTTGACAGACCTGGCTGCCAACACTCTCGCTTG 858
 QY 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
 Db 859 GCCAAGGCCAACACAGCACACCTCTGGAGGATATGACAGCCAGCTCACTCATTCACAGT 918
 QY 101 GlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 Db 919 CAGATGGAGAACATCACCACATATCTCTCAAGCCACGAGCAGAACCTGAAGACCTGCAG 978
 QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
 Db 979 GACTTACACAAAGATGCAGAGAATAGAACAGCCATCAAGTTCAACCAACTGAGGAAACGC 1038
 QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrrThrAlaHis 160
 Db 1039 TTCCAGCTTTTGAGACCGGATATTGTGAACATCATTAGCAATATCAGTTACACGCCAC 1098
 QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 Db 1099 CACCTGCGGACGCTGACCAGCAATCTAAATGAAGTCAGGACCCTTGACACATACCCCTT 1158
 QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 Db 1159 ACCAAACACACAGATGATCTGACCTCTCTGAATAATACCTGGCCCAACATCGTTGGAT 1218
 QY 201 SerValSerLeuArgMetGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
 Db 1219 TCTGTCTCTCAGGATGCACAAGATTGTGAGGTGAGGTTAGACACTGAAGTAGCC 1278
 QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
 Db 1279 AACTTATCAGTCATTATGGAAGAAATGAAGTAGTAGTCCCAAGCATGGTCAGCTCATC 1338
 QY 241 LysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGly 260
 Db 1339 AAGAAATTTTACAATACTACAAGGTCCACGGGCCCGGAGGTTCCAAGAGGTGACAGAGA 1398
 QY 261 SerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro 280
 Db 1399 TCCAGGAGACCCCTGGCCCAACTGGCAACAAGGAGACAGAAAGAGAGAGGGAGCCT 1458

```
QY 281 GlyProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGlu 300
Db 1459 GGACCACTGGGCTCGGGCTGAGAGAGGCCAAATGGACCACTGGTCCCGCGAGAG 1518
QY 301 ArgGlyGlySerGlySerGlySerGlnGlyProLysGlySerArgGlySerProGly 320
Db 1519 CGTGGCGGCAAGAGTCTAAAGGCTCCAGGGGCCCAAGGCTCCGTTCCCGTGGG 1578
QY 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProProGlyLysGlu 340
Db 1579 AAGCCGCGCCCTCAGGCGCCAGTGGGACCCAGGCGCCCGGGCCACAGCAAGAG 1638
QY 341 GlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
Db 1639 GGACTCCCGGGCCCTCAGGGCCCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGAG 1698
QY 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
Db 1699 CTTGGGGTGGTGGACCTCGGGACTGCCAGGCTTGCCCTGGGGTACAGGCATGCCAGGC 1758
QY 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
Db 1759 CCCAAGGGCCCCCGCCCTCTGGCCCATCAGGAGCGTGGTCCCTGGCCCTGCAG 1818
QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
Db 1819 AATGAGCCAAACCCGCGCAGGAGCAATGGCTGCCCGCCCTCACTGGAAGAACTTCACA 1878
QY 421 AspLysCysTrpTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
Db 1879 GACAAATGCTACTATTTTCAGTTGAGAAGAAATTTTGGAGATGCAAAAGCTTTTCTGT 1938
QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLys 460
Db 1939 GAAGACAAGCTTTCACATCTGTTTTCATAAACACTAGAGAGAACAGCAATGGATAA 1998
QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
Db 1999 AACAGATGGTAGGAGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAAT 2058
QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyLysPro 500
Db 2059 GAATGGAAGTGGCTGGATGGACATCTCCAGACTACAAAATGGAAAGCTGGACAGCCG 2118
QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
Db 2119 GATACTGGGGTCATGGCCATGGCCAGGAGAAAGCTGTGCTGGGTGATTTATGCTGGG 2178
QY 521 GlnTrpAsnAspPheGlnCysGluAspValAlaAsnAsnPheIleCysGluLysAspArgGlu 540
Db 2179 CAGTGGAGACGATTCCAATGTGAAGACGTCAATAACTTCATTTGCGAAAAAGACAGGGAG 2238
QY 541 ThrValLeuSerSerAlaLeu 547
Db 2239 ACAGTACTGTCATGCAATTA 2259
RESULT 3
AAH43036
ID AAH43036 standard; DNA; 2628 BP.
XX
AC AAH43036;
XX
DT 15-OCT-2001 (first entry)
XX
DE Nucleotide sequence of a human scavenger receptor.
KW Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
KW low density lipoprotein; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
```

```
FT CDS 74..2302
FT FT /*tag= a
XX /product= "scavenger receptor"
PN WO200159107-A1.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-JP00874.
PR 14-FEB-2000; 2000JP-0035155.
PR 10-OCT-2000; 2000JP-0309068.
XX
XX (FUSO ) FUSO PHARM IND LTD.
XX
XX Wakamiya N;
XX
XX WPI; 2001-497076/54.
XX P-PSDB; AAG63346.
XX
XX New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
XX useful for treatment and diagnosis of diseases associated with oxidized
XX low-density lipoprotein accumulation
XX
XX Claim 2; Page 79-84; 118pp; Japanese.
XX
XX The present sequence encodes a human scavenger receptor, designated
XX SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
XX are useful in clarifying the functions of macrophages and basal
XX immunity. They are also useful in the treatment, prevention, diagnosis
XX and investigation of diseases such as arteriosclerosis, diabetic
XX complications, bacterial infection and restenosis following angioplasty,
XX which are associated with accumulation of oxidized low density
XX lipoprotein and the binding of advanced glycation end-products into
XX cells.
XX
XX Sequence 2628 BP; 783 A; 661 C; 656 G; 528 T; 0 other;

Alignment Scores:
Pred. No.: 6,31e-96 Length: 2628
Score: 2969.00 Matches: 547
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-763-712A-2 (1-547) x AAH43036 (1-2628)
QY 1 MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuThrGlnValGln 20
Db 659 ATGTATTCTCATAATGTGGTTCATCATGAACCTCAACAACTGAACCTGACCGGTGCAG 718
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
Db 719 CAGAGGAACCTCATCAGGAATCTGCAGCGTCTGTGGATGACACAGCCAGCTATCCAG 778
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnGlnValPheLeuGlnAlaLysLysAspThr 60
Db 779 CGAATCAAGAACGACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGAGACAGC 838
QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
Db 839 GATTGGCTGAAGGAGAAAGTGCAGAGCTTGCAGACGCTGGCTGCCAACAACTCTCGGTTG 898
QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
Db 899 GCCAAGGCCCAACAGCACACCTTGGAGGATATGAACAGCCAGCTCAACTCATTTCAGGT 958
QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
Db 959 CAGATGGAGAACATCACCACCTATCTCTCAAGCCCAAGCAGACACCTGAAAGACCTGCAG 1018
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
```

Db 1019 GACTTACACAAAGATGCAGAGAAATAGAACAGCCATCAAGTTCAACCAACTGGAGAACGC 1078
Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleLeuSerAsnIleSerTyrThrAlaHis 160
Db 1079 TTCACAGCTCTTTGACACGGATATGTGAACTATTCATCAATATATCATGTTACACACCCAC 1138
Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
Db 1139 CACCTCGCGAGCGTGACACGCAATCAATGAAGTCAGGACCACTTGCACAGATACCCCTT 1198
Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
Db 1199 ACCAAACACACAGATGATCTGACCTCTTGAATAATACCTCGGCCAACATCCGTTTGAT 1258
Qy 201 SerValSerLeuArgMetGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
Db 1259 TCTGTTCTCTCAGAGATGCACACAGATTTGATGAGTCGAGGTAGACACTGAAGTAGCC 1318
Qy 221 AsnLeuSerValIleMetGluGlnMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
Db 1319 AACTTATCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATC 1378
Qy 241 LysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGly 260
Db 1379 AAGAAATTTTACAATACTACAAGGTCCACCGGGCCCGAGGGTCCCAAGAGGTGACAGAGA 1438
Qy 261 SerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro 280
Db 1439 TCCCAGGACCCCTCGGCCCACTGGCAACAGGAGACAGAAGGAGAGAGGGGAGGCT 1498
Qy 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlu 300
Db 1499 GGACCACTGGCCCTCGCGGTGAGAGAGGCCCAATGTGACCACTGGTCCCCCGGAGAG 1558
Qy 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
Db 1559 CGTGGCGCAAGAGGATCTAAAGGCTCCCAAGGCGCCCAAGGCTCCCGTGTCCCTCGG 1618
Qy 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
Db 1619 AAGCCCGCCCTCAGGGCCCACTGGGAGCCAGCCCGCCCGGCCCAAGGAGAGAG 1678
Qy 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
Db 1679 GGACTCCCGCCCTCAGGGCCCTCTGCTTCCAGGACITTCAGGGCACCTTGGGAG 1738
Qy 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyWetProGly 380
Db 1739 CCTGGGTGCTTGGACCTCGGGGAGTGCAGGCTTGCCTGGGGTACCAGGCATGCCAGGC 1798
Qy 381 ProLysGlyProProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
Db 1799 CCCAAGGGCCCCCGGCCCTCTGGCCCATCAGAGCGGTGGTCCCTCGCCCTGGCAG 1858
Qy 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
Db 1859 AATGAGCAACCCCGCACCGGAGACAATGGCTCCCGCCCTCACTGGAAGAACTTCACA 1918
Qy 421 AspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
Db 1919 GACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCCTTTCTGT 1978
Qy 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
Db 1979 GAAGACAGAGCTTTCACATCTTGTGTTTCAATAACACTAGAGAGACAGCAATGGATAAAA 2038
Qy 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
Db 2039 AAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCCCTCACAGACTCAGAGCGTGAAT 2098
Qy 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 500

Db 2099 GAATGGAAGTGGCTGGATGGGACATCTCAGACTTCAAAAAATTGGAAGCTGGACAGCGG 2158
Qy 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
Db 2159 GATAACTGGGTTCATGGCCATGGCCAGGAGAACTGTGCTGGTGTGATTATGCTGGG 2218
Qy 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
Db 2219 CAGTGAACGATTTTCCAATGTGAAGACGTCAATAACTTCATTTCGGAAAAAGACAGGAG 2278
Qy 541 ThrValLeuSerSerAlaLeu 547
Db 2279 ACAGTACTGTCTATCTGCATTA 2299
RESULT 4
ID AAI60628 standard; cDNA; 2181 BP.
XX AAI60628;
AC AAI60628;
XX 22-OCT-2001 (first entry)
DT Human polynucleotide SEQ ID NO 4617.
DE Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX Homo sapiens.
OS WO200153312-A1.
PN 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US34263.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-052317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR P-PSDB; AAM41472.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Claim 1; SEQ ID NO 4617; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 2181 BP; 664 A; 539 C; 535 G; 443 T; 0 other;

Alignment Scores:

Pred. No.: 8.86e-96 Length: 2181
 Score: 2963.00 Matches: 546
 Percent Similarity: 99.82% Conservative: 0
 Best Local Similarity: 99.82% Mismatches: 1
 Query Match: 99.80% Indels: 0
 DB: 22 Gaps: 0

US-09-763-712a-2 (1-547) x AAI60628 (1-2181)

QY 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
 DB 47 ATGTATTCTCATATATGGTTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAG 106
 QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
 DB 107 CAGAGGAACCTCATCAGGAATCTGCAGCGTCTGTGGATGACACAGCCAGGCTATCCAG 166
 QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnGlnValPheLeuGlnAlaLysLysAspThr 60
 DB 167 CGAATCAAGAACACACTTTCAAAATCTGCAGCAGGTTTTTCTCAAGCCCAAGAGGACAG 226
 QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
 DB 227 GATTGGCTGAAGAGAAAGTGCAGACTTGCAGACACTGGCTGCCAACCACTCTCGGTG 286
 QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
 DB 287 GCCAAAGCCCAACACACACCTCTGGAGGATATGAACAGCCAGCTCACTCACTACAGGT 346
 QY 101 GlnMetGluAsnIleThrPheSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 DB 347 CAGATGGAGAACATCACCACTATCTCTCAAGCCACAGCAGCAAACTGAAAGACCTGCG 406
 QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
 DB 407 GACTTACACAAAGATGCAGAGATAGAACAGCCATCAAGTTCAACCACTGGAGGAACCC 466
 QY 141 PheGlnLeuPheGluThrAspIleValAsnIleLeSerAsnIleSerThrAlaHis 160
 DB 467 TTCCAGCTCTTTGAGACGGATATTGTGAACATCATTAGCAATATCAGTTACACAGCCAC 526
 QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrCysThrAspThrLeu 180
 DB 527 CACCTGGCGACCGCTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGCACAGATACCT 586
 QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 DB 587 ACCAAACACACAGATGATCTGACCTCTTGAATATACCTGGCCAAACATCCGTTGGAT 646
 QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
 DB 647 TCTGTTTCTCTCAGGATGCACAAAGATTGATGAGTGCAGGTTAGACACTGAAGTAGCC 706
 QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
 DB 707 RACTTATCAGTGATTATCGAGAAATGAAGTAGTAGTACTCAAGCATGGTCAGCTCATC 766
 QY 241 LysAsnPheThrIleLeuGlnGlnProGlyProArgGlyProArgLysArgGly 260
 DB 767 AAGAAATTTTACAATACTACAAGGTCCACCGGCCCGCCAGGGTCCCAAGAGGTGACAGAGA 826
 QY 261 SerGlnGlyProGlyProThrClyAsnLysGlyGlnLysGlyGluLysGlyGluPro 280
 DB 827 TCCAGGAGACCCCTGGCCCCAAGTGGCAACAAAGGACAGAAAGAGAGAGAGGGGAGCCT 886

QY 281 GlyProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlu 300
 DB 887 GGACCACTGGCCCTCGGGTGAGAGAGGCCCAATTTGGACAGCTGCTCCCCCGGAGAG 946
 QY 301 ArgGlyGlyLysGlySerLysGlySerGlnGlnProLysGlySerArgGlySerProGly 320
 DB 947 CGTGGCGGCAAGAGATCTAAAGGCTCCAGGGCCCAAGGCTCCCGTGGTCCCTGGG 1006
 QY 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
 DB 1007 AAGCCCGGCCCTCAGGGCCCACTGGGGACCCAGGCCCCCGGCCCCAGCAGCAAGAG 1066
 QY 341 GlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
 DB 1067 GGACTCCCGGCCCTCAGGGCCCTCTGGCTTCCAGGGACTTCAGGGCCCGGTTGGGAG 1126
 QY 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
 DB 1127 CCTGGGTGCTGGACCTCGGGACTGCCAGGCTTGGCTGGGTACCGCATGCCAGGC 1186
 QY 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
 DB 1187 CCCAAGGGCCCGGCCCTCTGGCCCATCAGGAGCGGTGTGCCCTGGCCCTGTCAG 1246
 QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
 DB 1247 AATGAGCCAAACCCCGCAGCCGAGGACAAATAGCTGCCGCCCTCCTGGAAGAACTTCACA 1306
 QY 421 AspLysCysTyrrTyrrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
 DB 1307 GACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGTAGGATGCAAAAGCTTTCTGT 1366
 QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
 DB 1367 GAAGACAAGTCTTCACATCTTGTTCATAAACACTAGAGAGGAACAGCAATGGATAAAA 1426
 QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
 DB 1427 AAACAGATGGTAGGAGAGAGAGCCACTGGATCGCCCTCACAGACTCAGAGCGTCAAAAT 1486
 QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrrLysAsnTrpLysAlaGlyGlnPro 500
 DB 1487 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTCAAAAAATTTGAAAGCTGGACAGCG 1546
 QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrrAlaGly 520
 DB 1547 GATACTGGGTGATGGCCATGGCCAGGAGAGAGACTGTGCTGGTTGATTATGCTGGG 1606
 QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
 DB 1607 CAGTGGAAACGATTTCCCAATGTGAAGACGCTCAATAACTTCATTTGCGAAAAAGACAGGAG 1666
 QY 541 ThrValLeuSerSerAlaLeu 547
 DB 1667 ACAGTACTGTCTCTGCATTA 1687

RESULT 5

AAC66903

ID AAC66903 standard; cDNA; 2641 BP.

XX

XX AAC66903;

AC

XX 27-MAR-2001 (first entry)

DT

XX Human EXMAD-14 coding sequence SEQ ID NO: 39.

DE

XX Extracellular matrix and adhesion-associated protein; EXMAD; cancer;

KW inflammation; reproductive disorder; cardiovascular disorder;

KW immune disorder; musculoskeletal disorder; developmental disorder;

KW gastrointestinal disorder; cell proliferation disorder; ss.

XX

OS Homo sapiens.

XX WO200068380-A2.
 XX 16-NOV-2000.
 XX 10-MAY-2000; 2000WO-US12811.
 XX 11-MAY-1999; 99US-0133643.
 XX 23-AUG-1999; 99US-0150409.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DAM;
 PI Azimzai Y;
 DR WPI: 2001-007395/01.
 DR P-PSDB: AAB27236.
 XX
 PT Isolated polynucleotide encoding extracellular matrix or
 PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
 PT preventing disorders associated with expression of EXMAD such as
 PT proliferative, immune and genetic disorders -
 XX Claim 4; Page 121-122; 129pp; English.
 XX
 CC The present invention provides the protein and coding sequences for 25
 CC novel extracellular matrix and adhesion-associated proteins (EXMADs).
 CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
 CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
 CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
 CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
 CC useful in the prevention and treatment of cancers, cell proliferation,
 CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
 CC gastrointestinal disorders and inflammation.
 XX
 SQ Sequence 2641 BP; 823 A; 628 C; 608 G; 582 T; 0 other;

Alignment Scores:
 Pred. No.: 1,03e-95 Length: 2641
 Score: 2963.00 Matches: 546
 Percent Similarity: 99.82% Conservative: 0
 Best Local Similarity: 99.82% Mismatches: 1
 Query Match: 99.80% Indels: 0
 DB: 22 Gaps: 0

US-09-763-712A-2 (1-547) x AAC66903 (1-2641)

Qy 1 MetTyrSerHisAsnValIleMetAsnLeuAsnLeuAsnLeuThrGlnValGln 20
 Db 297 ATGTATTCTCATAATGTGGTTCATCATGAACCTCAACACCTGAACCTGACCCAGGTGCAG 356
 Qy 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
 Db 357 CAGAGGAACCTCATCAGCAATCTCGACGGGTCTGTGGATGACAAAGCCAGGCTATCCAG 416
 Qy 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
 Db 417 CGAATCAAGAACGACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCCAAGACGACAG 476
 Qy 61 AspIrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
 Db 477 GATTGGCTGGAAGGAAAGTGCAGAGCTTGACAGCGCTGGCTGCCCAACAACTCTCGCTTG 536
 Qy 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
 Db 537 GCCAAAGCCACACAGCACCTCGAGGATATGACACCCAGCTCACTCACTCACAGGT 596
 Qy 101 GlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 Db 597 CAGATGGGAACATCACCACTATCTCTCAAGCCACAGCAGACCACTGAAAGACCTGCAG 656
 Qy 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
 Db 1737 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTCAAAAAATTGAAAAAGCTGGACAGCG 1796

Db 657 GACTTACAAAGATGCAGAGATAGAACAGCCATCAAGTTCAACCACTGGAGGAGCGC 716
 Qy 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
 Db 717 TTCAGCTCTTTGAGACGGATATTGTGAACATCATTAATGCAATATCAGTTTACACGCCAC 776
 Qy 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 Db 777 CACCTGGGACGCTCACCCAGCAATCTAATGAAGTCAGGACCACCTTGCACAGATACCCCT 836
 Qy 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 Db 837 ACCAAACACAGATGATCTGACCTCTTGAATAATACCTGGCCCAACATCCGTTTGGAT 896
 Qy 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
 Db 897 TCTGTTCTCTCAGGATGCAACAGATTTGTAGGTGAGGTGAGCTAGACACTGAAGTAGCC 956
 Qy 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLeu 240
 Db 957 AACTTATCAGTGATATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATC 1016
 Qy 241 LysAsnPheThrIleLeuGlnGlnGlyProGlyProArgGlyProArgGlyAspArgGly 260
 Db 1017 AAGAAATTTTACAACTACTACAAGGTCCACCGGCCCCAGGGGTCCCAAGAGGTGACAGGA 1076
 Qy 261 SerGlnGlyProGlyProGlyProThrGlyAsnLysGlnLysGlyGluLysGlyGluPro 280
 Db 1077 TCCAGAGGACCCCTTGGCCCACTGGCAACAAAGGACAGAAAGGAGAGAGGGGGAGCCT 1136
 Qy 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlyGlu 300
 Db 1137 GGACCACTGCGCTCGGGTGAGAGAGGCCCAATTGGACAGCTGGTCCCCCGGAGAG 1196
 Qy 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
 Db 1197 CGTGGCGCAAGAGGATCTAAAGGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCCTGGG 1256
 Qy 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
 Db 1257 AAGCCCGCCCTCAGGGCCCGCAGTGGGACCCAGGCCCGCCCGGGCCCGCCAGGCAAGAG 1316
 Qy 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
 Db 1317 GGACTCCCGCCCTCAGGGCCCTCTGCTGCTCCAGGACTTCAGGGGACCCCTTGGGAG 1376
 Qy 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
 Db 1377 CCTGGGGTGCCTGGACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAGGCATGCCAGC 1436
 Qy 381 ProLysGlyProProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
 Db 1437 CCCAAGGGCCCCCGGCCCTCTCTGGCCCATCAGAGGGGTGGTCCCTGGCCCTGCAG 1496
 Qy 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
 Db 1497 AATGAGCCAAACCCCGCACCGGAGGACAATAGTCCCGCCCTCACTGGAAGAACTTCACA 1556
 Qy 421 AspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
 Db 1557 GACAAATGCTACTATTTTTTTCAGTTGAGAAAGAAATTTTTTGGAGGATGCAAAAGCTTCTGT 1616
 Qy 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
 Db 1617 GAAGNCAAGTCTTCATCTCTTGTTCATAAACACTAGAGAGGAACAGCAATGGATAAAA 1676
 Qy 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
 Db 1677 AAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCCCTCACAGACTCAGAGCGTGAATA 1736
 Qy 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 500
 Db 1737 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTCAAAAAATTGAAAAAGCTGGACAGCG 1796

QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
|||||
Db 1797 GATAACTGGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTGATTTATGCTGGG 1856

QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
|||||
Db 1857 CAGTGGAGGATTTCAATGTGAAGACGTCAATACTTCATTTGGCAAAAGACAGGGAG 1916

QY 541 ThrValLeuSerSerAlaLeu 547
|||||
Db 1917 ACAGTACTGTCAATCGCATTA 1937

RESULT 6
AAI58842
ID AAI58842 standard; cDNA; 2318 BP.
XX
AC AAI58842;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1045.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR P-PSDB: AAM39686.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 1045; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other;

Alignment Scores:

Pred. NO.: 1.63e-95 Length: 2318
Score: 2956.00 Matches: 547
Percent Similarity: 99.82% Conservative: 0
Best Local Similarity: 99.82% Mismatches: 1
Query Match: 99.56% Indels: 1
DB: 22 Gaps: 0

US-09-763-712a-2 (1-547) x AAI58842 (1-2318)

QY 1 MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
Db 47 ATGTATTCTCATAATGTGTCATCATGAACCTCAACAACCTGAACCTGACCCAGGTGCAG 106

QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
Db 107 CAGAGAACCTCATCCGAATCTGCAGCGTCTGTGGATGACACAGCCAGGCTATCCAG 166

QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
Db 167 CGAATCAAGAAGGACTTTCAAAATCTGCAGCAGGTTTTTCTTCAAGCCAAAGAGACAG 226

QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
Db 227 GATTGGCTGAAGGAGAAAGTGCAGAGCTTGCAGACACTGGCTGCCAACAACTCTCGCTTG 286

QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
Db 287 GCCAAAGCCACACAGCACCCCTGGAGGATATGAACAGCCACTCAACTCATTCACAGGT 346

QY 101 GlnMetGluAsnIleThrThrIleSerGln-AlaAsnGluGlnAsnLeuLysAspLeuG 120
Db 347 CAGATGGAGAACATCACCACATCTCTCAAGGCCAACGAGCAGAACCTGAAAGACCTGCA 406

QY 120 nAspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluGluAr 140
Db 407 GGACTTACACAAAGATGCAGAGATAGAACAGCCATCAAGTTCAACCAACTGGAGGAAG 466

QY 140 gPheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaH 160
Db 467 CTTCCAGCTCTTTGACAGCGGATATTGTGAACATCATATTAGCAATATCAGTTACAGAC 526

QY 160 sHisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLe 180
Db 527 CCACCTGCGGAGCTGACAGCAATCTAAATGAAGTCAGGACCACTTGACACAGATACCT 586

QY 180 uThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAs 200
Db 587 TACCAAAACACACAGATGATCTGACCTCTTGAATTAATACCTGGCCAACTCCGTTTGG 646

QY 200 pSerValSerLeuArgMetGlnAspLeuMetArgSerArgLeuAspThrGluValAl 220
Db 647 TTCTGTTTCTCTCAGAGTGCAACAGATTGTATGAGTTCGAGGTAGACACTGAAGTAGC 706

QY 220 aAsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuI 240
Db 707 CAACTTATCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGCTCAGCTCAT 766

QY 240 eLysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgG 260
Db 767 CAAGAATTTTCAATACTACAGGTCCACCGGGCCCCAGGGGTCCAGAGGTGACAGAGG 826

QY 260 ySerGlnGlyProGlyProThrThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPr 280
Db 827 ATCCAGGAGCCCTCTGCCCCACTGGCAACAGGACAGAAAGAGAGAGAGGGGGAGGCC 886

Qy 280 oGlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyG1 300
Dbb TGGACCACTGGCCCTGGGGTGGAGAGAGGCCAATTGGACCACTGGTCCCGCGGAGA 946
Qy 300 uArgGlyGlySerGlySerGlySerGlySerGlySerGlySerGlySerProG1 320
Dbb CGGTGGCGGCAAGAGATCTAAGAGCTCCAGGGCCCCAAGAGCTCCCGTGGTCCCGCTGG 1006
Qy 320 yLysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysG1 340
Dbb GAAGCCCGCCCTCAGGCGCCAGTGGGACCCAGGCGCCCGCGGCGCCACAGGCAAGA 1066
Qy 340 uGlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyG1 360
Dbb GGGACTCCCGGCGCTCAGGCGCTCCCTGGCTTCAGGGACTTCAGGGCACCGTGGGGA 1126
Qy 360 uProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProG1 380
Dbb GCTGGGGTGGCTGGACCTCGGGGACTCCAGGCTTGCCTGGGGTACCAGGCGATGCCAGG 1186
Qy 380 yProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuG1 400
Dbb CCCCAAGGCGCCCGCGCCCTCCTGGCCCATCAGGACGGTGGTGGCCCTGGCTGCA 1246
Qy 400 nAsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPhTh 420
Dbb GAATGAGCAACCCCGGCGCCAGGAGCAATGGCTGCCCGCTCACTGGAAGAACTTCAC 1306
Qy 420 rAspLysCysTrpTrpPheSerValGluLysLeuPheGluAspAlaLysLeuPheC 440
Dbb AGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTG 1366
Qy 440 sGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleL 460
Dbb TGAACACAAGCTTCCACATCTGTTTTCATAAACAACACTAGAGAGGAACAATGATGATAA 1426
Qy 460 sLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAs 480
Dbb AAAACAGATGTAGGAGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGCTGAAAA 1486
Qy 480 nGluTrpLysTrpLeuAspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnPr 500
Dbb CGAATGGAGTGGTGGTGGATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACACC 1546
Qy 500 oAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTrpAlaG1 520
Dbb GGATAACTGGGTGATGCCATGGCCAGGAGAGACTGTCTGGGTGATTATGCTGG 1606
Qy 520 yGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgG1 540
Dbb GCAGTGGGAACGATTTCCAATGTGAAGACGTCATAAATTCATTTCGAAAAAGACAGGA 1666
Qy 540 uThrValLeuSerSerAlaLeu 547
Dbb GACAGTACTGTCATCTGCATTA 1688
RESULT 7
ABQ92072
ID ABQ92072 standard; cDNA: 2929 BP.
XX AC ABQ92072;
XX AC ABQ92072;
DT 04-OCT-2002 (first entry)
XX Human polynucleotide SEQ ID NO 69.
DE Human; cytostatic; antirheumatic; antiarthritic; vulnery; analgesic;
KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW antitumor; fungicide; antidiabetic; antidiabetic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;

Alzheimer's disease; Parkinson's disease; Huntington's disease; osteoporosis; severe combined immunodeficiency; SCID; infection; multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.
Homo sapiens.
US2002065394-A1.
30-MAY-2002.
22-DEC-2000; 2000US-0745763.
18-MAR-1998; 98US-0040963.
(JACO/) JACOBS K.
(MCCO/) MCCOY J M.
(LAVA/) LAVALLIE E R.
(COLL/) COLLINS-RACIE L A.
(EVAN/) EVANS C.
(MERB/) MERBERG D.
(TREA/) TREACY M.
(SPAU/) SPAULDING V.
Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
Merberg D, Treacy M, Spaulding V;
P-PSDB; ABP61859.
WPI; 2002-582343/62.
Novel secreted or transmembrane protein and polynucleotide encoding the protein, useful for diagnosis and treatment of neurological disorders, cancer, autoimmune diseases, bone disorders and lung or liver fibrosis.

Claim 250; Page 231-232; 284pp; English.

The invention relates to human secreted or transmembrane protein (I), their fragments and is encoded by specific complementary deoxyribonucleic acid (cDNA) inserts (II), where the protein is substantially free from other mammalian proteins. (I) are useful for preventing, treating or ameliorating a medical condition, especially immunological treatment or prevention of tumors. (I) exhibits activity relating to angiogenesis, cytokine, cell proliferation, cell differentiation, antiinflammatory, stem cell growth factor activity and activin or inhibin-related activities. (I) can be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. (I) induces the proliferation of neural cells and regeneration of nerve and brain tissue and is useful for the treatment of central and peripheral nervous system diseases and neuropathies, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis. (I) is involved in chemotactic or chemokinetic activity, regulation of haematopoiesis and is useful for treating myeloid or lymphoid cell disorders, platelet disorders such as thrombocytopaenia and for regeneration of bone, cartilage, tendon, ligament and/or nerve tissue growth and in tissue repair, healing of burns, incisions, ulcers, for treating osteoporosis, osteoarthritis, bone degenerative disorders or periodontal disease. (I) is also useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, various immune deficiencies and disorders including severe combined immunodeficiency (SCID), bacterial or fungal infections, autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic reactions and conditions such as asthma or other respiratory problems. (II) is useful to express recombinant protein, as markers for tissues in which the corresponding protein is preferentially expressed and in gene therapy. The present sequence is that of a polynucleotide of the invention.

Sequence 2929 BP; 901 A; 699 C; 692 G; 637 T; 0 other;

Alignment Scores:

| | | | |
|------------|----------|----------|------|
| Pred. No.: | 9,74e-95 | Length: | 2929 |
| Score: | 2936.00 | Matches: | 545 |

Percent Similarity: 99.63% Conservative: 0
Best Local Similarity: 99.63% Mismatches: 2
Query Match: 98.89% Indels: 1
DB: 24 Gaps: 0

US-09-763-712a-2 (1-547) x ABQ92072 (1-2929)

```
QY 1 MetTyrSerHisAsnValIleMetAsnLeuAsnLeuValGln 20
|||||
DB 652 ATGATTCTCATTAATGTGTCATCATGAATC-AACAACCTGNAACCTGACCCAGGTGCAG 710
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
|||||
DB 711 CAGAGGAACCTCATCAGGAATCGCAGCGTCTGTGGATGACACAAGCCAGGCTATCCAG 770
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysAspThr 60
|||||
DB 771 CGAATCAAGAAGACATTTCAAAATCTGCAGCAGGTTTCTTCAAGCCAAGGACAGC 830
QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
|||||
DB 831 GATTGCTGAGGAGAAAGTCAGAGCTTGCAGACGCTGGCTGCCAACAACCTCGCGTTG 890
QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
|||||
DB 891 GCCAAAGCCAAACAACGACACCTCGGAGGATATGAACAGCAGCTCAACTCATTCACAGGT 950
QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
|||||
DB 951 CAGATGGAGAACATCACCACTATCTCTCAAGCCACGAGCAGAACCTGAAAGACCTGCGAG 1010
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
|||||
DB 1011 GACTTACAAAGATGCAGAGATAGAACAGCCATCAAGTTCAACCACTGGAGGAACGC 1070
QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerThrAlaHis 160
|||||
DB 1071 TTCCAGCTCTTTGAGACGGATATTGTGAACATCATTTAGCAATATCATAGTTACACAGCCAC 1130
QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
|||||
DB 1131 CACCTCGGACGCTGACAGCAATCTTAATGAAGTCAGACCACTTGACAGATACCCCTT 1190
QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
|||||
DB 1191 ACCAAACACACAGATGATCTGACCTCCTTGAATAATACCTTGCCCAACATCCGTTTGGAT 1250
QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
|||||
DB 1251 TCTGTTTCTCTCAGGATGCAACAAGATTTGATGAGTCTGAGGTTAGACACTGAAGTAGCC 1310
QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
|||||
DB 1311 AACTTATCAGTATTATGAAGAATGAAGTAGTAGACTCCAAGCATGGTCAGCTCATC 1370
QY 241 LysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGly 260
|||||
DB 1371 AAGAATTTTACATACTACAAGGTCCACGGGGCCCCAGGGGTCCAAGAGGTGCACAGAGA 1430
QY 261 SerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyLysGlyGluPro 280
|||||
DB 1431 TCCAGGGGACCCCTCGGCCCACTTGGCAACAAGGGACACAAAGGAGAGAGGGGGAGCCT 1490
QY 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlu 300
|||||
DB 1491 GGACCACTGGCCCTCGGGTGAGAGAGGCCCAATTTGGACCACTGGTCCCCCGGAGAG 1550
QY 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
|||||
DB 1551 CGTGGCGGCAAGGATCTAAAGCTCCCAAGGCCCCCAAGGCTCCCGTGGTCCCCCTGGG 1610
QY 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyLysGlu 340
|||||
```

```
DB 1611 AAGCCCGGCCCTCAGGGCCCCAGTGGGGACCCAGGCCCCCCGGGCCACAGCAAGAG 1670
QY 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
|||||
DB 1671 GGACTCCCCGGCCCTCAGGGCCCTCTGCTTCCAGGGACTTCAGGGCCACCGTTGGGGAG 1730
QY 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
|||||
DB 1731 CCTGGGGTCCCTGGGACTCGGGACTCCAGGCTTCCCTGGGGTACCAGGCATGCCAGGC 1790
QY 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
|||||
DB 1791 CCCAAGGGCCCCCGGCCCTCTGCGCCATCAGGAGCGTGTGTCCTGGCCCTGCAG 1850
QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
|||||
DB 1851 AATGACCAACCCCGCCAGGAGCAATAGCTGCCGCCCTCCTGGAAGAACTTTCACA 1910
QY 421 AspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
|||||
DB 1911 GACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAGCTTTTCTGT 1970
QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
|||||
DB 1971 GAAGACAAGCTCTTCACATCTTGTTTTTCATAAACACTAGAGAGAACCAATGGATAAAA 2030
QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
|||||
DB 2031 AACAGATGTTAGGAGAGAGAGCCACTGGATCGGCTCAGAGCTCAGAGCGTGAAT 2090
QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 500
|||||
DB 2091 GAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGGAAGCTGGACAGCCG 2150
QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
|||||
DB 2151 GATACTGGGGTCATGGCCATGGCCAGGAGAAAGACTGCTGGGTGTGATTTATGCTGGG 2210
QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
|||||
DB 2211 CAGTGAACGATTTCCAATGTGAAGACGTCATAACTTCATTTGCCAAAAAGACAGGGAG 2270
QY 541 ThrValLeuSerSerAlaLeu 547
|||||
DB 2271 ACAGTACTGTCATCTGCATTA 2291
RESULT 8
AAV5746
ID AAV55746 standard; cDNA; 2930 BP.
AC AAV55746;
XX
DT 23-MAR-1999 (first entry)
XX
DE Human secreted protein clone bv227_1 coding sequence.
XX
KW Secreted protein; human; nutritional supplements; immune stimulant;
KW immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis;
KW activin/inhibin; chemokinesis; haemostasis; thrombolysis;
KW receptor/ligand activity; anti-inflammatory; tumour inhibitor;
KW cadherin/tumour invasion suppressor; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 67..693
FT /tag= a
XX
PN WO9855614-A2.
XX
PD 10-DEC-1998.
XX
PF 01-JUN-1998; 98WO-US11210.
```


XX 29-MAY-1998; 98US-0087255.
 PR 04-JUN-1997; 97US-0868696.
 PR 04-JUN-1997; 97US-0868697.
 PR 04-JUN-1997; 97US-0868698.
 PR 04-JUN-1997; 97US-0868898.
 PR 04-JUN-1997; 97US-0868899.
 PR 04-JUN-1997; 97US-0868900.
 PR 04-JUN-1997; 97US-0869191.
 PR 04-JUN-1997; 97US-0869192.
 PR 04-JUN-1997; 97US-0869193.
 PR 04-JUN-1997; 97US-0869194.
 XX (GEMY) GENETICS INST INC.
 XX Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER;
 PI McCoy JM, Racie LA, Spaulding V, Treacy M;
 XX WPI: 1999-059912/05.
 DR P-PSDB; AAW73628.
 XX New polynucleotides encoding secreted human proteins - derived from
 PT human foetal brain, adult brain, foetal kidney, adult ovary, adult
 PT retina, adult placenta or adult uterus cDNA libraries
 XX Claim 26; Page 87-88; 127pp; English.
 XX This sequence encodes a human secreted protein of the invention.
 CC This DNA sequence was isolated from a human adult brain cDNA
 CC library, and was designated clone bv227_1. The DNAs and proteins
 CC are predicted to have biological activities which would make them
 CC suitable for treating, preventing or ameliorating medical conditions in
 CC humans and animals, although no supporting data is given. Suggested
 CC activities include nutritional sources or supplements, immune
 CC stimulating or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity. The DNAs
 CC are also stated to be useful for gene therapy. A host cell transfected
 CC with the DNA, or its subfragments and variants is useful for recombinant
 CC production of the human secreted protein clones.
 XX SQ Sequence 2930 BP; 902 A; 699 C; 692 G; 637 T; 0 other;

Alignment Scores:
 Pred. No.: 9,75e-95 Length: 2930
 Score: 2936.00 Matches: 545
 Percent Similarity: 99.63% Conservative: 0
 Best Local Similarity: 99.63% Mismatches: 2
 Query Match: 98.89% Indels: 1
 DB: 20 Gaps: 0

US-09-763-712A-2 (1-547) x AAV55746 (1-2930)

QY 1 MetTyrSerHisAsnValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
 DB 652 ATGTATTCTCAATGTTGGTCATCATGAACTC-AACAACCTGAACCTGACCAGGTGCAG 710
 QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
 DB 711 CAGAGGAACCTCATCACCAATTCGAGCGGTCTGTGGATGACACAAGCCAGGCTATCCAG 770
 QY 41 ArgIleLysAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysAspThr 60
 DB 771 CGAATCAAGACGACTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGACACG 830
 QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAsnAsnSerAlaLeu 80
 DB 831 GATTGGCTGAAGGAGAAAGTCAGAGCTTGCAGAGCTGGCTGGCTGCCAACAACCTCTCGGTG 890
 QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100

DB 891 GCCAAAGCCAAACAGCACACCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGT 950
 QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
 DB 951 CAGATGGAGAACATCACCACATATCTCTCAAGCCAGCAGACAACTGAAAGACCTGCAG 1010
 QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluGluArg 140
 DB 1011 GACTTACACAAAGATGCAGAGAATAGACAGCCATCAAGTTCAACCAACTGGAGGAACGC 1070
 QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerIleThrAlaHis 160
 DB 1071 TTCCAGCTCTTTGAGACGGATATTGTGAACATCATAGCAATATCAGTTACACAGCCAC 1130
 QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 DB 1131 CACCTGGGCGCTGACCAGCAATCTAATGAAGTCAGGACCCTTGCACACATACCCCT 1190
 QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 DB 1191 ACCAAACACACAGATGATCTGACCTCTTGAATAATACCTGCACCAATCCGTTTGAT 1250
 QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
 DB 1251 TCTGTTCTCTCAGGATGCAACAAGATTTGATGAGGTCGAGGTAGACACTGAAGTAGCC 1310
 QY 221 AsnLeuSerValIleMetGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
 DB 1311 AACTTATCAGTCATTTATGGAAGAAATGAAGCTAGTAGACTCCCAAGCATGGTCAGCTATC 1370
 QY 241 LysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyProArgGlyAspArgGly 260
 DB 1371 AAGAAATTTTACAATACTACAAGTCCACCGGCCCCAGGGGTCCAAGAGGTGACAGAGGA 1430
 QY 261 SerGlnGlyProGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro 280
 DB 1431 TCCCAAGGACCCCTTGGCCCACTGGCAACAAGGACAGAGAGAGAGAGAGAGAGAGAGCCT 1490
 QY 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlyGlu 300
 DB 1491 GGACCACTGGCCCTGCGGGTGCAGAGAGCCCAATGGACAGCTGGTCCCCCGGAGAG 1550
 QY 301 ArgGlyGlyLysSerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
 DB 1551 CGTGGCGCAAGAGATCTAAAGGCTCCCAAGGCCCAAGGCTCCCGTGGTTCCTCCCTGG 1610
 QY 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
 DB 1611 AAGCCCGGCCCTCAGGGCCCAAGTGGGACCCAGGCCCGCCCGGCCCAAGAGAG 1670
 QY 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
 DB 1671 GGACTCCCGGCCCTCAGGGCCCTCTGCTTCCAGGGACTTCAGGGACCCGTGGGGAG 1730
 QY 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
 DB 1731 CCTGGGTGCTTGCAGCTCGGGGACTGCCAGGCTTCCCTGGGTACCGCATGCCAGC 1790
 QY 381 ProLysGlyProGlyProGlyProGlyProSerGlyAlaValProLeuAlaLeuGln 400
 DB 1791 CCAAGAGGCCCCCGGCCCTCTCTGCCCATCAGGAGCGGTGGTGGCCCTGCCCTGCAG 1850
 QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
 DB 1851 AATGAGCAACCCCGCCAGCGGAGCAATACTGCCGCCCTCCTCTGGAAGAACTTCACA 1910
 QY 421 AspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
 DB 1911 GACAAATGCTACTATTTTCACTTGAGAAGAAATTTTGGAGATGCAAGACTTCTCTGT 1970
 QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
 DB 1971 GAAGACAAGTCTTCACATCTTGTTCATAAAACACTAGAGAGAACAGCAATGGATAAA 2030

QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
Db 2031 AAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAAT 2090
QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 500
Db 2091 GAATGGAAGTGGCTGGTGGACATCTCCAGACTACAAAAATTGAAAGCTGGACAGCG 2150
QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
Db 2151 GATACTGGGTGATGGCCATGGCCAGGAGAGACTGCTGGGTGATTTATGCTGGG 2210
QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPhelleCysGluLysAspArgGlu 540
Db 2211 CAGTGAACGATTTCATCAAGTGAAGACGTCAATACTTCATTTGCCAAAAAGACAGGAG 2270
QY 541 ThrValLeuSerSerAlaLeu 547
Db 2271 ACAGTACTGTCACTGCATTA 2291
RESULT 9
AAH43037
ID AAH43037 standard; cDNA; 2637 BP.
XX
AC AAH43037;
XX
DT 15-OCT-2001 (first entry)
XX
DE Nucleotide sequence of a human scavenger receptor.
XX
KW Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
KW low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT 92..2320
FT /*tag= a
FT /product= "scavenger receptor"
XX
XX WO200159107-A1.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-JP00874.
XX
PR 14-FEB-2000; 2000JP-0035155.
PR 10-OCT-2000; 2000JP-0309068.
XX
XX (FUSO) FUSO PHARM IND LTD.
XX
PA
XX Wakamiya N;
PI
PI
DR WPI: 2001-497076/54.
DR P-PSDB; AAG63347.
XX
XX New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
PT useful for treatment and diagnosis of diseases associated with oxidized
PT low-density lipoprotein accumulation
XX
PS Claim 6; Page 88-93; 118pp; Japanese.
XX
CC The present sequence encodes a human scavenger receptor, designated
CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
CC are useful in clarifying the functions of macrophages and basal
CC immunity. They are also useful in the treatment, prevention, diagnosis
CC and investigation of diseases such as arteriosclerosis, diabetic
CC complications, bacterial infection and restenosis following angioplasty,
CC which are associated with accumulation of oxidized low density
CC lipoprotein and the binding of advanced glycation end-products into
CC cells.

XX
SQ Sequence 2637 BP; 829 A; 617 C; 635 G; 556 T; 0 other;
Alignment Scores:
Pred. No.: 8,44e-89 Length: 2637
Score: 2765.00 Matches: 502
Percent Similarity: 96.71% Conservative: 27
Best Local Similarity: 91.77% Mismatches: 18
Query Match: 93.13% Indels: 0
DB: 22 Gaps: 0
US-09-763-712a-2 (1-547) x AAH43037 (1-2637)
QY 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
Db 677 ATGTATTCTCAGAGCGTGGTTATCATGAACCTCAACAACCTGAACCTTAACCCAGGTTCCAG 736
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspAspThrSerGlnAlaIleGln 40
Db 737 CAGAGGAACCTTATCTCAAAATCTGCAGACGTCTGTGGATGACACAAAGCCCTGGCCATCCAG 796
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
Db 797 CGAATTAAGAATGATTTCAAAATCTGCAGCAGGTTTCTTCAACCCAAAGAGACACC 856
QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
Db 857 GATTGGCTAAAGGAAAAAGTACAGAGCTTGCGACACATTGGCTGCCAACCAACTCTGCCCTG 916
QY 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
Db 917 GCCAAAGCCCAACAATGACACCCCTAGAGGATATGAATAGCCAGCTCAGCTCATTTACAGGT 976
QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
Db 977 CAGATGGCAACATTACCACTATCTCACAGGCCAACAGCAGCAGGCTGAAAGACCTTCAG 1036
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluGluArg 140
Db 1037 GACTTACACAAGGATACAGAAAAATAGAACAGCTGTCAAGTTACGCCAACTTGAGGAACGC 1096
QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
Db 1097 TTCCAGGTCTTTGAGACAGATATTGTGAACATCATTTAGCAACATCAGCTACACAGCCCAT 1156
QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
Db 1157 CACCTGAGGACACTGACAGCAATCTGATGATGTTAGGACACACATGCACAGACCTTG 1216
QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
Db 1217 ACCAGACACAGGATGACCTGACCTCCTCTTGAATAACACACTAGTCAACATCCCGTTGGAT 1276
QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
Db 1277 TCTATTTCTCTCAGGATGCAGCAAGACATGATGAGTCAAAAGTTAGACTAGGTTGGCC 1336
QY 221 AsnLeuSerValIleMetGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
Db 1337 AACTTATCAGTGGTTATGGAAGAGATGAACCTGGTTGACTCCCAAGCAGCGTCACTCATC 1396
QY 241 LysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGly 260
Db 1397 AAGAACTTTTACCATTCTACAAGGTCTCTCTGGCCCCAGAGGTCCAAAGAGTGACAGAGGA 1456
QY 261 SerGlnGlyProProGlyProThrGlyAsnLysGlyGlnLysGlyGlyGluPro 280
Db 1457 TCTCAGGGACCACTGGTCCAACTGGCAACAAGGGGACAGAAAGGAGAGAGGAGAGCCT 1516
QY 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlyGlu 300
Db 1517 GGTCCACCTTGGCCCTCGGGGTGAGGGGGGACAAATTTGGACAGTGGCCCTCTCTGGAGAG 1576

```
Qy 301 ArgGlyGlyLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
Db 1577 CGTGCAGCAAAAGGATCAAAAGGCTCACAGGTCCTCAAGGATCTCGTGGTCCCGAGG 1636
Qy 321 LysProGlyProGlnGlyProSerGlyAspProGlyProProGlyProGlyLysGlu 340
Db 1637 AAGCCTGGCCCTCAAGGACCTAGTGGGACCCAGGACCAAGTCCACGAGCAAGGAT 1696
Qy 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
Db 1697 GGACTCCCTGGCCCTCAGGCGCTCTCGCTCCAGGACCTACAGGCACTGTGGGTGAG 1756
Qy 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
Db 1757 CCTGGAGTACCTGGACCTCGGGGGTGGCAGGCTTGCAGGGGTGCCAGGCACTGGCTGGG 1816
Qy 381 ProLysGlyProGlyProGlyProGlyProSerGlyAlaValProLeuAlaLeuGln 400
Db 1817 CCTAAGGACCACTTGGCCCTCAGGCGCTTGCAGGCAATGGAGCACTTGGCTTGCAG 1876
Qy 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
Db 1877 ATGAACCAACCCAGCATCAGAGTCAACGGATGTCGCCCTCCTCGGAGACTTCACA 1936
Qy 421 AspLysCysTyrTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
Db 1937 GATAAATGCTACTATTTTCATTTGAAAAAGAAATTTTGAAGATGCTAAGCTTTTCTGT 1996
Qy 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLys 460
Db 1997 GAAGACAAATCTCCCATCTCGTTTTCATAACTCAAGAGAAAGACACCAATGGATAAA 2056
Qy 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
Db 2057 AAGCATACCGTGGGAGAGAAAGCCATGGATCGGCCCTCAGACTCAGACAGAGGAGC 2116
Qy 481 GluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 500
Db 2117 GAATGGAAGTGGCTAGACGGGTCACTGTGTTGATTACAAAACTGGAAGCTGGACAACCA 2176
Qy 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGly 520
Db 2177 GATAACTGGGCGAGTGGCCATGGCCAGGAGAAAGACTGTGCTGGCTTCAATTTACGACGA 2236
Qy 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
Db 2237 CAGTGGATGACTTCCAGTGTGATGAATCAATACTTCACTTTGTGAGAGAGGAGGAG 2296
Qy 541 ThrValLeuSerSerAlaLeu 547
Db 2297 GCAGTACCATCATCATATTA 2317
RESULT 10
ABL95574
ID ABL95574 standard; cDNA; 2005 BP.
XX
AC ABL95574;
XX
XX Human angiotensin related cDNA PRO7223 SEQ ID NO: 27.
DT 19-JUL-2002 (first entry)
XX
DE Human; angiotensin; PRO protein; cardiovascularisation; wound; cancer;
KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
KW cardiast; cytostatic; antiangiogenic; hypotensive; vulnerary;
KW antihypertensive; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200208284-A2.
XX
XX 31-JAN-2002.
PD
XX
```

```
PF 09-JUL-2001; 2001WO-US21735.
XX 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220664P.
PR 28-JUL-2000; 2000WO-US20710.
PR 02-AUG-2000; 2000US-222695P.
PR 17-AUG-2000; 2000US-064357.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 07-SEP-2000; 2000US-230978P.
PR 15-SEP-2000; 2000US-000000P.
PR 18-SEP-2000; 2000US-0664610.
PR 18-SEP-2000; 2000US-0665350.
PR 24-OCT-2000; 2000US-242922P.
PR 08-NOV-2000; 2000US-0709238.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 22-JAN-2001; 2001US-0767609.
PR 28-FEB-2001; 2001US-0796498.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001US-0506666.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 25-MAY-2001; 2001WO-US17092.
PR 30-MAY-2001; 2001US-0870574.
PR 30-MAY-2001; 2001WO-US17443.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 28-JUN-2001; 2001WO-US00000.
XX (GETH ) GENENTECH INC.
PA (BAKE) BAKER K P.
PA (FERR) FERRARA N.
PA (GERB) GERBER H.
PA (GERR) GERRITSEN M E.
PA (GODD) GODDARD A.
PA (GODO) GODOWSKI P J.
PA (GURN) GURNEY A L.
PA (HILL) HILLAN K J.
PA (MARS) MARSTERS S A.
PA (PANJ) PAN J.
PA (PAON) PAONI N F.
PA (STEP) STEPHAN J F.
PA (WATA) WATANABE C K.
PA (WILL) WILLIAMS P M.
PA (WOOD) WOOD W I.
XX
XX Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
PI Stephan JF, Watanabe CK, Williams PM, Wood WL, Ye W;
XX
XX WPI; 2002-171999/22.
DR P-PSDB; ABB95436.
XX
XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or angiogenic disorders in a mammal -
XX
XX Claim 1; Fig 27; 567pp; English.
PS
XX The present invention provides the protein and coding sequences of human
CC PRO proteins. These are useful for treating or diagnosing a
CC cardiovascular, endothelial or angiogenic disorder, including cardiac
```

CC hypertrophy, trauma, cancer, age-related macular degeneration,
CC atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound
CC healing. The present sequence is a coding sequence of the invention.
XX
SQ Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;

Alignment Scores: 3.35e-86 Length: 2005
Pred. No.: 2688.00 Matches: 495
Score: 2688.00
Percent Similarity: 99.60% Conservatives: 2
Best Local Similarity: 99.20% Mismatches: 2
Query Match: 90.54% Indels: 1
DB: 24 Gaps: 0

US-09-763-712a-2 (1-547) x ABL95574 (1-2005)

QY 49 LeuGlnGlnValPheLeuGlnAlaLysLysAspThrAspTrpLeuLysGluLysValGln 68
Db : : : : :
Db 8 GTCGGCAGGTTTCTTCAAGCCAAAGAGGACACGGATTGGCTGAAGGAGAAAGTGCAG 67
QY 69 SerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeuAlaLysAlaAsnAsnAspThrLeu 88
Db : : : : :
Db 68 AGCTTCGACACACTGGCTGCCAACAACTCTGGCTGGCCAAAGCCAAACACGACACCCCTG 127
QY 89 GluAspMetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrIle 108
Db : : : : :
Db 128 GAGGATATGAACAGCCAGCTCAACTATTACAGAGTCAGATGGAGAACATCACCCTATC 187
QY 109 SerGlnAlaAsnGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsn 128
Db : : : : :
Db 188 TCTCAAGCCAAACGAGAGAACCTGAAAGACCTGCAGGACTTACACAAAGATGCAGAGAAT 247
QY 129 ArgThrAlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIle 148
Db : : : : :
Db 248 AGAACAGCCATCAAGTTCAACCACTGGAGGACGCTTCCAGCTCTTTGAGACGGATATT 307
QY 149 ValAsnIleIleSerAsnIleSerThrAlaHisHisLeuArgThrLeuThrSerAsn 168
Db : : : : :
Db 308 GTGAACATCATTAGCAATATCAGTTACACAGCCACCACTGCGGACGCTGACCAACAT 367
QY 169 LeuAsnGluValArgThrThrCysThrAspThrLeuThrLysHisThrAspLeuThr 188
Db : : : : :
Db 368 CTAAATGAAGTCAGGACCACTTGCACAGATACCTT-ACCATAACACACAGATCTGACC 426
QY 189 SerLeuAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGln 208
Db : : : : :
Db 427 TCCTTGAATATACCTGGCCCAACATCCGTTGGATTCTGTTCTCTCAGGATGCCAA 486
QY 209 AspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGlu 228
Db : : : : :
Db 487 GATTTCATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGAAGAA 546
QY 229 MetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGly 248
Db : : : : :
Db 547 ATGAAGCTAGTAGCTCAAGCATGGTCAAGTCACTCATCAAGATTTTACAATACTACAAGT 606
QY 249 ProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThr 268
Db : : : : :
Db 607 CCACCGGGCCCCAGGGTCCAGAGGTGACAGAGGATCCAGGGACCCCTGGCCCAACT 666
QY 269 GlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGlu 288
Db : : : : :
Db 667 GGCAACAAGGGACAGAAAGGAGAGGGGAGCGCTGGACCACTGGCCCTCGGGGTGAG 726
QY 289 ArgGlyProIleGlyProIleGlyProProGlyGluArgGlyGlyLysGlySerLysGly 308
Db : : : : :
Db 727 AGAGGCCCAATTTGGACCAAGCTGGTCCCGCGGAGAGCGGTGGCGGCAAGGATCTAAAGC 786
QY 309 SerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSer 328
Db : : : : :
Db 787 TCCAGGGCCCCAAAGGCTCCCGTGGTCCCTTGGAAAGCCCGGCTCAGGGCCCCCAGT 846

QY 329 GlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyPro 348
Db : : : : :
Db 847 GGGAGCCAGGCCCGCGGGCCACCAGGCAAGAGAGGACTCCCGGCCCTCAGGGCCCT 906
QY 349 ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGly 368
Db : : : : :
Db 907 CCTGGCTTCAGGGACTTCAGGGCACCGTTGGGGAGCCTGGGGTGCCTGACCTCGGGGA 966
QY 369 LeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProPro 388
Db : : : : :
Db 967 CTGCCAGGCTTGCTGGGTACCAGGATGCCAGGGCCCCAAGGGCCCCCGGCCCTCCCT 1026
QY 389 GlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGlu 408
Db : : : : :
Db 1027 GGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAAGAGGAGGAGGAGGAG 1086
QY 409 AspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerVal 428
Db : : : : :
Db 1087 GACAATAGCTGCCCGCTCAGTGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTT 1146
QY 429 GluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerHisLeuVal 448
Db : : : : :
Db 1147 GGAAGAATAATTTTGAGGATGCAAGCTTCTCTGAAGACAAGTCTTCACATCTGTGT 1206
QY 449 PheIleAsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSer 468
Db : : : : :
Db 1207 TTCATAAACACATAGAGAGGAACAGCAATGGATAAAAAACAGATGTTAGGAGAGAGC 1266
QY 469 HisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThr 488
Db : : : : :
Db 1267 CACTGGATCGGCTCACAGACTCAGACGCTGAAATGAAATGAAAGTGGGTGATGGGACA 1326
QY 489 SerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGly 508
Db : : : : :
Db 1327 TCTCCAGACTACAAANTTGAAGAGCTGGACCCGGTAACCTGGGGTGTATGCCCATGGG 1386
QY 509 ProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGlu 528
Db : : : : :
Db 1387 CCAGGAGAAGACTGTGCTGGTGTGATTATGCTGGCAGTGGAACCATTTTCCAATGTGA 1446
QY 529 AspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 547
Db : : : : :
Db 1447 GACGTCAATAACTTCATTTGCGAAAAAGACAGGAGACAGTACTGTCTATCTGCATTA 1503
RESULT 11
ABL88085
ID ABL88085 standard; cDNA; 2005 BP.
XX
AC ABL88085;
XX
DT 16-MAY-2002 (first entry)
XX
DE Human PROT223 cDNA sequence SEQ ID NO:27.
XX
KW Human; angiogenesis; cardiant; cytostatic; antiangiogenic; hypotensive;
KW vulnerable; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;
KW anglogenetic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
KW age-related macular degeneration; arterial restenosis; angina;
KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;
KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
KW wound healing; chromosome mapping; gene mapping; gene; ss.
OS Homo sapiens.
XX
XX WO200200690-A2.
PN
XX
PD 03-JAN-2002.
XX
XX 20-JUN-2001; 2001WO-US19692.
PF
XX
XX 23-JUN-2000; 2000US-213637P.
PR

| | | | |
|--|-----|---|------|
| DB: | 24 | Gaps: | 0 |
| US-09-763-712A-2 (1-547) x ABL88085 (1-2005) | | | |
| Qy | 49 | LeuGlnGlnValPheLeuGlnAlaLysLysAspThrAspTrpLeuLysGluLysValGln | 68 |
| Db | 8 | GTCCGGCAGGCTTTTCCTCAAGCCACGAAGGACACGGATTGGCTGAAGGAAAGTCAG | 67 |
| Qy | 69 | SerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeuAlaLysAlaAsnAsnAspThrLeu | 88 |
| Db | 68 | AGCTTGCAGACACTGGCTGCCAACAACTCTGCTTGCCCAAGCCAAACAGCACCCCTG | 127 |
| Qy | 89 | GluAspMetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIle | 108 |
| Db | 128 | GAGCATATGAACACCCAGCTCAACTCATTCACAGCTCAGATGGAGAAACATCACCACTATC | 187 |
| Qy | 109 | SerGlnAlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsn | 128 |
| Db | 188 | TCTCAAGCCACAGCAGACACCTGTAAGAGACTCGACAGCTTACACAAAGATGCAGAGAAT | 247 |
| Qy | 129 | ArgThrAlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIle | 148 |
| Db | 248 | AGAACAGCATCAAGTTCACCAACTCGAGGAACGCTTCAGACTCTTTGAGACGGATATT | 307 |
| Qy | 149 | ValAsnIleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsn | 168 |
| Db | 308 | GTGAACATCATTAGCAATATCAGTTACACAGCCACCACCTGGGACGCTGACCAGCAAT | 367 |
| Qy | 169 | LeuAsnGluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThr | 188 |
| Db | 368 | CTAATCAAGGTCAGGACCACTTGCACAGATACCTT-ACCAACACACAGATGATCTGACC | 426 |
| Qy | 189 | SerLeuAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGln | 208 |
| Db | 427 | TCCTTGAATTAATACCTTGGCCACATCCGTTTGGATTCTGTGTTCTCTCAGAGTGCAACAA | 486 |
| Qy | 209 | AspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGlu | 228 |
| Db | 487 | GATTTGATGAGTCCGAGTTAGACACTGAAGTACCCAACTTATCAGTGATTATGGAAGAA | 546 |
| Qy | 229 | MetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGly | 248 |
| Db | 547 | ATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAAATTTTACAATACTACAAGT | 606 |
| Qy | 249 | ProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThr | 268 |
| Db | 607 | CCACCGGCCCAAGGCTCCAGAGGTCACAGAGGATCCACAGGACCCCTTGCCCAACT | 666 |
| Qy | 269 | GlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyProAlaGlyGlu | 288 |
| Db | 667 | GGCAACAGGCACAGAAGGAGAGAGAGGGGAGCCTGGACCACCTGGCCCTGCGGGTGAG | 726 |
| Qy | 289 | ArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyLysGlyLysLysGly | 308 |
| Db | 727 | AGAGGCCCAATTTGACCAGCTGGTCCCGCGGAGAGCGTGGCGGCAAAAGGATCTAAAGGC | 786 |
| Qy | 309 | SerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSer | 328 |
| Db | 787 | TCCCAGGCCCAAGGCTCCCGTGGTTCCCTTGGGAAGCCCGGCCCTCAGAGGCCCACT | 846 |
| Qy | 329 | GlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyPro | 348 |
| Db | 847 | GGGACCCAGGCCCCCGGCCACACAGGCAAGAGGAGGATCTCCCGGCCCTCAGAGGCCCT | 906 |
| Qy | 349 | ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGly | 368 |
| Db | 907 | CCTGGCTTCCAGGACATTCAGGCCACCGCTTGGGAGAGCCTGGGTGCTTGGACCTCGGGGA | 966 |
| Qy | 369 | LeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProPro | 388 |
| Db | 967 | CTGCCAGGCTTGCTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGGCCCTTCCT | 1027 |
| Qy | 389 | GlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGlu | 408 |

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 3.35e-86 | Length: | 2005 |
| Score: | 2688.00 | Matches: | 495 |
| Percent Similarity: | 99.60% | Conservative: | 2 |
| Best Local Similarity: | 99.20% | Mismatches: | 2 |
| Query Match: | 90.54% | Indels: | 1 |

```

Db 1027 GGCCCATCAGAGCGTGGTCCCTCGCCCTCAGCAATGAGCAACCCCGGACCGGAG 1086
QY 409 AspAsnGlyCysProHisTriPlysAsnPhenThrAspLysCysTyrTrpPheSerVal 428
Db 1087 GACAATAGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTCAGTT 1146
QY 429 GluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerHisLeuVal 448
Db 1147 GAGAAAGAAATTTTGGAGTGCAGAAAGCTTTCTGTGAAGACAAAGCTTTCACATCTGTT 1206
QY 449 PheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSer 468
Db 1207 TTCATAAACACTAGAGAGGAACAGCATGATGATAAAAAACAGATGCTAGGAGAGAGAGC 1266
QY 469 HisTriPleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLysTrpLeuAspGlyThr 488
Db 1267 CACTGGATCGGCTCACAGACTCAGAGCGTGAANAATGAATGGAAGTGGCTGGATGGGACA 1326
QY 489 SerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGly 508
Db 1327 TCTCCAGACTACAAAATTTGAAAGCTGGAGACGCCGATAAATGGGTCTATGCCCATGGG 1386
QY 509 ProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGlu 528
Db 1387 CCAGGAGAAAGACTGTGCTGGGTTGATTTATGCTGGCAGTGGAAACGATTTCCAAATGTGA 1446
QY 529 AspValAsnAsnPhenIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 547
Db 1447 GAGTCNTAACTTCATTTCCGAAAAGACAGGGAGACAGTACTGTCTATCTGCATTATTA 1503

RESULT 12
AAH43054
ID AAH43054 standard; DNA; 2256 BP.
XX
AC AAH43054;
XX
XX 15-OCT-2001 (first entry)
XX
DE Nucleotide sequence of a human scavenger receptor.
XX
KW Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
KW low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 74..1936
FT FT /*tag= a
FT FT /product= "scavenger receptor"
XX
XX WO200159107-A1.
XX
PD 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-JP00874.
XX
XX 14-FEB-2000; 2000JP-0035155.
XX
XX 10-OCT-2000; 2000JP-0309068.
XX
XX (FUSO ) FUSO PHARM IND LTD.
XX
XX Wakamiya N;
XX
XX WPI; 2001-497076/54.
XX
XX P-PSDB; AAG63350.
XX
XX New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
XX useful for treatment and diagnosis of diseases associated with oxidized
XX low-density lipoprotein accumulation
XX
XX Claim 4; Page 105-109; 118pp; Japanese.

```

```

XX
CC The present sequence encodes a human scavenger receptor, designated
CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
CC are useful in clarifying the functions of macrophages and basal
CC immunity. They are also useful in the treatment, prevention, diagnosis
CC and investigation of diseases such as arteriosclerosis, diabetic
CC complications, bacterial infection and restenosis following angioplasty,
CC which are associated with accumulation of oxidized low density
CC lipoprotein and the binding of advanced glycation end-products into
CC cells.
XX
SQ Sequence 2256 BP; 721 A; 524 C; 528 G; 483 T; 0 other;
Alignment Scores:
Pred. No.: 1.15e-68 Length: 2256
Score: 2187.00 Matches: 423
Percent Similarity: 77.33% Conservat: 0
Best Local Similarity: 77.33% Mismatches: 0
Query Match: 73.66% Indels: 124
DB: 22 Gaps: 1
US-09-763-712A-2 (1-547) x AAH43054 (1-2256)
QY 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnAsnLeuThrGlnValGln 20
Db 659 ATGTATTCTCATATGTGGTGCATCATCAACCAACCTGAACCTGACCCAGGTGCAG 718
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspAspThrSerGlnAlaIleGln 40
Db 719 CAGAGGAACCTCATCAGCAATCTGCAGCGGTCTGTGTGATGACACAGCCAGGCTATCCAG 778
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnGlnValPheLeuGlnAlaLysLysAspThr 60
Db 779 CGAATCAAGAACGACATTCATAAATCTGCAGCAGGTTTTTCTCAAGCCAAGAGGACAG 838
QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaIleAsnAsnSerAlaLeu 80
Db 839 GATTGGCTGAAGGAGAAAGTGCAGAGCTTGCAGCAGCTGGCTGCCAACAACTCTGCGTTG 898
QY 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
Db 899 GCCAAAGCCCAACCAACGACACCTGGAGGATATGAACAGCCAGCTCACTCATTCACAGGT 958
QY 101 GlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
Db 959 CAGATGGAGAACATCACCACCTATCTCTCAAGCCAACGAGCAGAACCTGGAAGACCTGCAG 1018
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
Db 1019 GACTTACAAAGATGCAGAGAAATAGAACCCATCAAGTTCAACCAACTGGAGGAACGC 1078
QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 160
Db 1079 TTCAGCTCTTTGAGACGGATATTGTGAACATCATTTAGCAATATCATGTTACACAGCCAC 1138
QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
Db 1139 CACCTGGGACGCTGACCAGCAATCTAAATGAAGTCAAGGACCACTTGCACAGATACCTT 1198
QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
Db 1199 ACCAAACACACAGATGATCTGACCTCTTGAATAATACCTTGGCCCAACATCCGTTGGAT 1258
QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuValAla 220
Db 1259 TCTGTTTCTCTCAGGATGCACCAAGATTTGTAGGTTCGAGGTAGACACTGAAGTAGCC 1318
QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
Db 1319 AACTTATCAGTGATTATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGTGTAGCTCATC 1378
QY 241 LysAsnPheThrIleLeuGlnGlyProArgGlyProArgGlyAspArgGly 260

```

Db 1379 AAGAAATTTTACAACTACAAAGTCCACCGGGCCCGGGGTCCCAAGAGGTGCACAGAGGA 1438
 Qy 261 SerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro 280
 Db 1439 TCCAGAGGACCCCTGGCCCACTGGCAACAGGAGCAGAAAGAGAGAGAGGGAGCCT 1498
 Qy 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGlu 300
 Db 1499 GGACCACTGGCCCTGGC----- 1516
 Qy 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 320
 Db 1516 ----- 1516
 Qy 321 LysProGlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGlu 340
 Db 1516 ----- 1516
 Qy 341 GlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
 Db 1516 ----- 1516
 Qy 361 ProGlyValProGlyProArgGlyLeuProGlyValProGlyMetProGly 380
 Db 1516 ----- 1516
 Qy 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValProLeuAlaLeuGln 400
 Db 1516 ----- 1516
 Qy 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPhetr 420
 Db 1517 -----GGCTGCCCGCCCTCACTGGAAGAACTTCACA 1546
 Qy 421 AspLysCysTrpTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
 Db 1547 GACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGTGCAGAACTTTCTGT 1606
 Qy 441 GluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLys 460
 Db 1607 GAAGACAAGTCTTCACATCTTTTTCATAAACACTAGAGAGGAACAGCAATGGATAAA 1666
 Qy 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
 Db 1667 AAACAGATGTPAGGAGAGAGAGCACTGGATCGCCCTCAGACTCAGAGCGTGAAT 1726
 Qy 481 GluTrpLysTrpLeuAspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnPro 500
 Db 1727 GAATGGAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACGCG 1786
 Qy 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTrpAlaGly 520
 Db 1787 GATAACTGGGTGTCATGGCCATGGCCAGGAGAGACTGTGCTGGTGTGATTATGCTGG 1846
 Qy 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
 Db 1847 CAGTGGACGATTTCCATGTGAAGACGCTCAATACTTCATTTCGGAAGAGAGAGGAG 1906
 Qy 541 ThrValLeuSerSerAlaLeu 547
 Db 1907 ACAGTACTGTCTATGTCATTA 1927
 RESULT 13
 ID AAS71133 standard; cDNA; 1521 BP.
 XX
 AC AAS71133;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #6937.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

food supplement; medical imaging; diagnostic; genetic disorder; ss.
 Homo sapiens.
 WO200175967-A2.
 11-OCT-2001.
 30-MAR-2001; 2001WO-US08631.
 31-MAR-2000; 2000US-0540217.
 23-AUG-2000; 2000US-0649167.
 (HYSE-) HYSEQ INC.
 Drmanac RT, Liu C, Tang YT;
 WPT; 2001-639362/73.
 P-PSDB; ABG06946.
 New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits and to assess
 biodiversity
 Claim 1; SEQ ID No 6937; 103pp; English.
 The invention relates to isolated polynucleotide (I) and
 polypeptide (II) sequences. (I) is useful as hybridisation probes,
 polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 and gene mapping, and in recombinant production of (II). The
 polynucleotides are also used in diagnostics as expressed sequence tags
 for identifying expressed genes. (I) is useful in gene therapy techniques
 to restore normal activity of (II) or to treat disease states involving
 (II). (II) is useful for generating antibodies against it, detecting or
 quantitating a polypeptide in tissue, as molecular weight markers and as
 a food supplement. (II) and its binding partners are useful in medical
 imaging of sites expressing (II). (I) and (II) are useful for treating
 disorders involving aberrant protein expression or biological activity.
 The polypeptide and polynucleotide sequences have applications in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits to assess biodiversity
 and to produce other types of data and products dependent on DNA and
 amino acid sequences. AAS64197-AAS94564 represent novel human
 diagnostic coding sequences of the invention.
 Note: The sequence data for this patent did not appear in the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 1521 BP; 386 A; 420 C; 425 G; 290 T; 0 other;
 Alignment Scores:
 Pred No.: 2.5e-52 Length: 1521
 Score: 1715.50 Matches: 298
 Percent Similarity: 94.43% Conservative: 7
 Best Local Similarity: 92.26% Mismatches: 7
 Query Match: 57.78% Indels: 11
 DB: 23 Gaps: 1
 US-09-763-712A-2 (1-547) x AAS71133 (1-1521)
 Qy 222 LeuSerValIleMetGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLys 241
 Db 130 CTACATGTCCTCATTTATGCAATGAGAAATTCACCGCACAGG----- 171
 Qy 242 AsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySer 261
 Db 172 -----CAAGGTCCACCGGGCCCGGGTCCAAAGAGGTGACAGAGATCC 216
 Qy 262 GlnGlyProProGlyProThrGlyAsnLysGlyGlnLysGlyGlyGluProGly 281
 Db 217 CAGGAGACCCCTGGCCCACTGGCAACAGGAGAGAGAGAGAGAGAGGAGGAGCCTGGA 276

QY 282 ProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArg 301
 Db 277 CCACCTGGCCCTCGGGTCAGAGAGGCCCAATTGGACAGCTGGTCCCCCGAGAGCGT 336
 QY 302 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 321
 Db 337 GCGCGCAAGAGATCTAAAGGCTCCAGGGGCCCAAGGCTCCGCTGGTTCCTCCCTGGGAG 396
 QY 322 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGly 341
 Db 397 CCGGCGCCCTCAGGGCTCCAGTGGGACCCAGGCCCCCGGGCCACAGGCAAGAGGGA 456
 QY 342 LeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluPro 361
 Db 457 CTCGCGGCGCCCTCAGGGCCCTCTGGCTTCAGAGGACTTCAGGGCACCGTTGGGGAGCCT 516
 QY 362 GlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyPro 381
 Db 517 GGGGTGCTGGACCTCGGGGACTGCGAGGCTTGCCTGGGGTACCAGGCATGCGAGGCCCC 576
 QY 382 LysGlyProGlyProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsn 401
 Db 577 AAGGCCCCCGGCGCTCTGGCCATCAGAGGCGGGTGGTCCCTGCGCTGCAGAGAT 636
 QY 402 GluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAsp 421
 Db 637 GAGCAACCCAGCAGCGGAGGACAATGGCTGCGCGCTCACTGGAAGAATTCACAGAC 696
 QY 422 LysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGlu 441
 Db 697 AAATGCTACTATTTTCAAGTTCAGAAAGAAATTTTGAGGATGCAAGGCTTTCTGTGAA 756
 QY 442 AspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLys 461
 Db 757 GACAGCTCTTCACATCTGTGTTTCATTAACACTAGAGGAGACACCAATGGTAAATAAA 816
 QY 462 GlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGlu 481
 Db 817 CAGATGGTAGGAGAGAGAGGACCTGGATGGCCCTCACAGACTCAGAGCGTGAAAAATGAA 876
 QY 482 TrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAsp 501
 Db 877 TGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATGGAAAGCTGGAGCGCGGAT 936
 QY 502 AsnTrpGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGln 521
 Db 937 AACTGGGTCATGGCCATGGGCCAGGAGAGACTGTGCTGGGTTGATTTATCTGGGCAG 996
 QY 522 TrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThr 541
 Db 997 TGAACGATTTCCAAATGTGAAGACGTCAATAACTTCAATTTGGAAAAAGACAGGGAGACA 1056
 QY 542 ValLeuSer 544
 Db 1057 GTCTTCAGC 1065
 RESULT 14
 AAX27858
 ID AAX27858 standard; DNA; 3685 BP.
 XX
 AC AAX27858;
 XX
 DT 02-JUN-1999 (first entry)
 XX
 DE Human CSR3 protein coding sequence.
 XX
 KW Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage;
 KW scavenger receptor protein; intracellular stress; arteriosclerosis;
 KW diabetic circulatory obstruction; microbial infection; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9909159-A1.

XX 25-FEB-1999.
 PD
 XX 12-AUG-1998; 98WO-JP03602.
 PF
 XX 30-JUL-1998; 98JP-0230121.
 PR
 PR 13-AUG-1997; 97JP-0233396.
 XX
 PA (NLSB) JAPAN TOBACCO INC.
 XX
 PI Nakamura Y, Tokino T;
 DR WPI; 1999-181032/15.
 DR P-PSDB; AAY00994.
 XX
 PT Scavenger receptor proteins - for treatment and diagnosis of
 PT disorders involving cell stress
 XX
 PS Claim 6; Page 142-150; 175pp; Japanese.
 XX
 CC This sequence encodes the human cellular stress response 3 (CSR3) protein
 CC of the invention. The CSR proteins are macrophage scavenger receptor
 CC proteins. The CSR proteins can be used in the treatment, gene therapy
 CC and diagnosis of diseases in which intracellular stress is important,
 CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial
 CC infection. Expression of the proteins is induced in vivo in response to
 CC intracellular stress, and inhibits cell death as a result of such stress.
 XX
 SQ Sequence 3685 BP; 901 A; 1142 C; 972 G; 669 T; 1 other;
 Alignment Scores:
 Pred. No.: 7,76e-19 Length: 3685
 Score: 765.00 Matches: 154
 Percent Similarity: 60.41% Conservative: 81
 Best Local Similarity: 39.59% Mismatches: 134
 Query Match: 25.77% Indels: 20
 DB: 20 Gaps: 4
 US-09-763-712a-2 (1-547) x AAX27858 (1-3685)
 QY 11 LeuAsnAsnLeuAsnLeuThrGlnValGlnGlnArgAsnLeuIleThrAsnLeuGlnArg 30
 Db 808 GTGCACACAGATCAACTTCCCGTGGGCGAGACTTCCGAGTGGATCCACGGATCCAGCGG 867
 QY 31 SerValAspThrSerGlnAlaIleGlnArgIleLysAsnAspPheGlnAsnLeuGln 50
 Db 868 AAGACAGAGAGAGAGACCTGACCTCCAGAGAGATTGTCCACGACTGGCAGAACTACACA 927
 QY 51 GlnValPheLeuGlnAlaLysLysAspThrAspTrpLeuLysGluLysValGlnSerLeu 70
 Db 928 CGGCTCTTCAGCGGCGCTGCGCACCACTCCACCAAGACTGGAGAGCGGCTCAAGAACATC 987
 QY 71 Gln---ThrLeuAlaAlaAsnAsnSerAlaLeuAlaLysAlaAsnAspThrLeuGlu 89
 Db 988 CAGGCCACCCCTGGGGCGCTCTCCAGCGCATCAGCCAG---AACTCAGAGAGCATGCAC 1044
 QY 90 AspMetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSer 109
 Db 1045 GACCTGGTACTCCAGTCCATGGCTTGCAGCTGCAGCTGGATAACATCTCGTCTTCTG 1104
 QY 110 GlnAlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGlnAsnArg 129
 Db 1105 GATGACACCAAGAGAACAATGCATGATCTTACAGTACCACCTACCCACTACGCCCAAGACGC 1164
 QY 130 ThrAlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleVal 149
 Db 1165 ACTGTGGAGAGGTTTGAGTCTCTGGAAGGACCATGGCTTCTCAGCAGAGATTGAATGGC 1224
 QY 150 AsnIleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeu 169
 Db 1225 ACCATCTTCCACCAACATCAATGCCACCGACCAACACCGTGCACAGCATGCTCAAGTACCTG 1284
 QY 170 AsnGluValArgThrThrCysThrAspThrLeuThrLysHisThrAspLeuThrSer 189


```
Db 1285 GATGAGCTCGGCTCTCTCCGACCTGGGCTTCCACACCCATCGCGAGGCTCTACTAC 1344
QY 190 LeuAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAsp 209
Db 1345 CTGACAAAGTCTCTCCATCATGCTGGCCACACAGACCTCTCGGAGCGCTTCAGC 1404
QY 210 LeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMet 229
Db 1405 CTGCTCAGTGGCCGCTGGACCTCAACGTCGCGAACCCTCTCCATGATGCTGGAGGAGATG 1464
QY 230 LysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyPro 249
Db 1465 AAGGAGTGGACACACAGCATGAGAAATCTTCGCAATGTCCACCATCTCAGAGGTGCC 1524
QY 250 ProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGly 269
Db 1525 CCGCGCCCTCCAGGACCAAGAGATCAAGGAGATATGGGGCTGAAAGGGCTGTGGC 1584
QY 270 AsnLysGlyGlnLysGlyGlu-----Lys 277
Db 1585 GGCAGAGGCCCGAAAGGAGACCCGCGCATCTTGGGCCCTGGGACCCCGAGGTCCTCAG 1644
QY 278 GlyGluProGlyProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyPro 297
Db 1645 GGGCAACCTGGAGAGCGCGGCTCTGGGAGAAAGGGCCCTGTGGCCCTCGAGGGTTC 1704
QY 298 ProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGly 317
Db 1705 CCAGGCGCTCAAGGCTCAAGGGGAGCTTGGAACTGGGGCCGAGG----- 1752
QY 318 SerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyProProGlyProPro 337
Db 1753 -----GGACAGCCAGGCCCAAGGGGACATAGGGCCCGCCAGGCGCAGAGGGCCCGG 1806
QY 338 GlyLysGluGlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThr 357
Db 1807 GGGTCTCCAGGGCCCTCAGGGGCTCAGGGAAACCGGGAATTCGAGGGAAGACAGGGTCA 1866
QY 358 ValGlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly 377
Db 1867 CCAGGCGAGCGGGGCCATGGGGCTTAGGGTGAACAGGATCCAGGGTCCCGCTGTGT 1926
QY 378 MetProGlyProLysGlyProProGly 386
Db 1927 CTCGGGGGCTTCCAGGTCCACAGGA 1953

RESULT 15
AAAX27856
ID AAAX27856 standard; DNA; 3810 BP.
XX
AC AAAX27856;
XX
DT 02-JUN-1999 (first entry)
XX
DE Human CSRL1 protein coding sequence.
XX
KW Cellular stress response protein; CSRL1; CSRL2; CSRL3; human; macrophage;
KW scavenger receptor protein; intracellular stress; arteriosclerosis;
KW diabetic circulatory obstruction; microbial infection; ss.
XX
OS Homo sapiens.
XX
PN W09509159-A1.
XX
PD 25-FEB-1999.
XX
PF 12-AUG-1998; 98WO-JP03602.
XX
PR 30-JUL-1998; 98JP-0230121.
XX
PR 13-AUG-1997; 97JP-0233396.
XX
PA (NTSB ) JAPAN TOBACCO INC.
```

```
XX Nakamura Y, Tokino T;
XX WPI: 1999-181032/15.
DR P-PSDB; AAY00992.
XX
PT Scavenger receptor proteins.- for treatment and diagnosis of
PT disorders involving cell stress
XX
PS Claim 4; Page 119-127; 175pp; Japanese.
XX
CC This sequence encodes the human cellular stress response 1 (CSRL) protein
CC of the invention. The CSR proteins are macrophage scavenger receptor
CC proteins. The CSR proteins can be used in the treatment, gene therapy
CC and diagnosis of diseases in which intracellular stress is important,
CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial
CC infection. Expression of the proteins is induced in vivo in response to
CC intracellular stress, and inhibits cell death as a result of such stress.
XX
SQ Sequence 3810 BP; 892 A; 1213 C; 1049 G; 655 T; 1 other;

Alignment Scores:
Pred. No.: 7,96e-19 Length: 3810
Score: 765.00 Matches: 154
Percent Similarity: 60.41% Conservative: 81
Best Local Similarity: 39.59% Mismatches: 134
Query Match: 25.77% Indels: 20
DB: Gaps: 4

US-09-763-712A-2 (1-547) x AAX27856 (1-3810)
QY 11 LeuAsnAsnLeuAsnLeuThrGlnValGlnGlnArgAsnLeuIleThrAsnLeuGlnArg 30
Db 933 GTGCACCAGCTCACTTCCCGTGGGGCAGACTTCCGAGTGGATCCACGGATCCAGCG 992
QY 31 SerValAspAspThrSerGlnAlaIleGlnArgIleLysAsnAspPheGlnAsnLeuGln 50
Db 993 AAGACAGACGAGGAGACCTGACCTCCAGAGAGATTGTCCACGACTGGCAGACTACACA 1052
QY 51 GlnValPheLeuGlnAlaLysLysAspThrAspTrpLeuLysGlnLysValGlnSerLeu 70
Db 1053 CGGCTCTTCGCGGCTTCCGCGACCCACTCCACAGACTGGAGAGCGGCTCAAGAACATC 1112
QY 71 Gln---ThrLeuAlaAlaAsnAsnSerAlaLeuAlaLysAlaAsnAspThrLeuGlu 89
Db 1113 CAGGCCACCTGGGGCTCTCCACAGCGCATCAGCGAG---AAGTCAGAGACATGCAC 1169
QY 90 AspMetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrIleSer 109
Db 1170 GACCTGGTACTCCAGGTCATGGGCTTGCAGCTGCAGCTGGATGAACATCTCGTCTCTG 1229
QY 110 GlnAlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArg 129
Db 1230 GATGACACGAGAGAACATGATGATCTCAGTACCATACCCACTACGCCAGAACCCG 1289
QY 130 ThrAlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleVal 149
Db 1290 ACTGTGGAGAGTTTCACTCTCTGGAGAGCGCATGGCTTCTCAGAGATTGAATGGC 1349
QY 150 AsnIleIleSerAsnIleSerTyThrAlaHisHisLeuArgThrLeuThrSerAsnLeu 169
Db 1350 ACCATCTTCCCAACATCAATGCCACCGACACACCGTGCACAGCATGCTCAAGTACCTG 1409
QY 170 AsnGluValArgThrThrCysThrAspThrLeuThrLysHisThrAspLeuThrSer 189
Db 1410 GATGAGTGGGCTCTCTCCGACCTGGGCTTCCACACCCATCGCGAGGCTCTACTAC 1469
QY 190 LeuAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAsp 209
Db 1470 CTGACAAAGTCTCTCCATCATGCTGGGACACACAGACTGCTCGGGAGCGCTTCAGC 1529
QY 210 LeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMet 229
Db 1584 GATGAGTGGGCTCTCTCCGACCTGGGCTTCCACACCCATCGCGAGGCTCTACTAC 1649
```


GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:26:21 ; Search time 58.1137 Seconds
(without alignments)
2886.622 Million cell updates/sec

Title: us-09-763-712a-2
Perfect score: 2969
Sequence: 1 MYSHVYIMNLNLNFQV.....EDVNNFICKDRETVLSSAL 547

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPRO_pool/US09763712/runat_14032003_100951_18187/app_query.fasta_1.1877
-DB=Issued_Patents_NA -QMT=fastcap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712 -CGN_1_108 @runat_14032003_100951_18187 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MWAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | ID | Description |
|------------|-------|-------|-------|--------|------------------|-------------------|
| 1 | 459.5 | 15.5 | 810 | 1 | US-08-642-255-60 | Sequence 60, Appl |
| 2 | 452.5 | 15.2 | 3181 | 1 | US-08-655-086-1 | Sequence 1, Appl |
| 3 | 448.5 | 15.1 | 756 | 1 | US-08-642-255-50 | Sequence 50, Appl |
| 4 | 436.5 | 14.7 | 1608 | 4 | US-09-029-348-19 | Sequence 19, Appl |
| 5 | 428 | 14.4 | 4031 | 1 | US-08-159-784-1 | Sequence 1, Appl |
| 6 | 424.5 | 14.3 | 1572 | 4 | US-09-297-269-39 | Sequence 39, Appl |
| 7 | 419 | 14.1 | 2543 | 1 | US-08-555-669-11 | Sequence 11, Appl |
| 8 | 419 | 14.1 | 2543 | 3 | US-09-073-663-11 | Sequence 11, Appl |
| 9 | 412.5 | 13.9 | 432 | 1 | US-08-642-255-48 | Sequence 4, Appl |
| 10 | 409 | 13.8 | 3394 | 1 | US-08-159-784-4 | Sequence 4, Appl |
| 11 | 402.5 | 13.6 | 1416 | 1 | US-07-621-091G-1 | Sequence 1, Appl |
| 12 | 402.5 | 13.6 | 1416 | 2 | US-08-399-889-1 | Sequence 1, Appl |

| | | | | | | |
|------|-------|------|-------|---|--------------------|-------------------|
| 13 | 402.5 | 13.6 | 1416 | 3 | US-09-167-364-1 | Sequence 1, Appl |
| 14 | 402.5 | 13.6 | 1416 | 4 | US-09-439-897-1 | Sequence 1, Appl |
| 15 | 400.5 | 13.5 | 4359 | 4 | US-09-484-970B-4 | Sequence 4, Appl |
| 16 | 399 | 13.4 | 1881 | 4 | US-09-029-348-20 | Sequence 20, Appl |
| 17 | 394.5 | 13.3 | 1560 | 2 | US-08-794-795-5 | Sequence 5, Appl |
| 18 | 394.5 | 13.3 | 1560 | 4 | US-09-249-200-5 | Sequence 5, Appl |
| 19 | 394 | 13.3 | 5102 | 1 | US-08-494-168-1 | Sequence 1, Appl |
| 20 | 393.5 | 13.3 | 503 | 4 | US-09-297-269-40 | Sequence 40, Appl |
| 21 | 391.5 | 13.2 | 61663 | 4 | US-09-453-702B-264 | Sequence 264, App |
| 22 | 391.5 | 13.2 | 61663 | 4 | US-09-453-702B-62 | Sequence 62, Appl |
| 23 | 389.5 | 13.1 | 1703 | 2 | US-08-794-795-1 | Sequence 1, Appl |
| 24 | 389.5 | 13.1 | 1703 | 4 | US-09-249-200-1 | Sequence 1, Appl |
| 25 | 380.5 | 12.8 | 1868 | 1 | US-08-392-367B-1 | Sequence 1, Appl |
| 26 | 380.5 | 12.8 | 1868 | 3 | US-08-893-467A-1 | Sequence 1, Appl |
| 27 | 356 | 12.0 | 1839 | 1 | US-08-383-744-1 | Sequence 1, Appl |
| 28 | 356 | 12.0 | 1839 | 2 | US-08-999-336-1 | Sequence 1, Appl |
| 29 | 356 | 12.0 | 1839 | 5 | PCT-US96-01427-1 | Sequence 1, Appl |
| 30 | 355.5 | 12.0 | 2383 | 4 | US-09-523-487-9 | Sequence 9, Appl |
| 31 | 355.5 | 12.0 | 2409 | 3 | US-09-320-095-9 | Sequence 9, Appl |
| 32 | 350 | 11.8 | 1458 | 4 | US-09-111-470-3 | Sequence 3, Appl |
| C 33 | 343 | 11.6 | 506 | 1 | US-08-253-155A-61 | Sequence 61, Appl |
| C 34 | 331.5 | 11.2 | 2824 | 2 | US-09-010-928B-3 | Sequence 3, Appl |
| C 35 | 330.5 | 11.1 | 392 | 4 | US-09-404-879A-372 | Sequence 372, App |
| C 36 | 329 | 11.1 | 9827 | 4 | US-09-453-702B-56 | Sequence 66, Appl |
| C 37 | 327 | 11.0 | 48908 | 4 | US-09-453-702B-137 | Sequence 137, App |
| C 38 | 325 | 10.9 | 45175 | 4 | US-09-453-702B-116 | Sequence 116, App |
| C 39 | 323.5 | 10.9 | 3404 | 4 | US-09-453-702B-94 | Sequence 94, Appl |
| C 40 | 322 | 10.8 | 1370 | 4 | US-09-111-470-9 | Sequence 9, Appl |
| C 41 | 315.5 | 10.6 | 2851 | 4 | US-09-535-521-1 | Sequence 1, Appl |
| C 42 | 315.5 | 10.6 | 2851 | 4 | US-09-535-521-3 | Sequence 3, Appl |
| C 43 | 313 | 10.5 | 2028 | 1 | US-08-453-117-1 | Sequence 1, Appl |
| C 44 | 313 | 10.5 | 2028 | 2 | US-08-948-222-1 | Sequence 1, Appl |
| C 45 | 313 | 10.5 | 2028 | 2 | US-08-973-145-1 | Sequence 1, Appl |

ALIGNMENTS

RESULT 1
US-08-642-255-60
; Sequence 60, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph A.
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 810 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic"
US-08-642-255-60

```

| | |
|------------------------|----------|
| Alignment Scores: | |
| Pred. No.: | 6.52e-18 |
| Score: | 459.50 |
| Percent Similarity: | 53.16% |
| Best Local Similarity: | 48.95% |
| Query Match: | 15.48% |
| DB: | 1 |
| Length: | 810 |
| Matches: | 93 |
| Conservative: | 8 |
| Mismatches: | 66 |
| Indels: | 23 |
| Gaps: | 4 |

US-09-763-712A-2 (1-547) x US-08-642-255-60 (1-810)

| | | | |
|----|-----|--|-----|
| QY | 247 | GlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGly | 266 |
| Db | 52 | AAAGGTGCTCCGGAGCTCGAGCCACCGGGTAGCGTGGGATCCGGACACCGGGT | 111 |
| QY | 267 | ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProProGlyProProAla | 286 |
| Db | 112 | GCACCTGGCCACCGGGTCGGCTGGATCTAGAGGTGACCCAGGACCGCTGGTCTCGG | 171 |
| QY | 287 | GlyGluArgGlyPro-----TleGlyPro | 294 |
| Db | 172 | GGACCTGCAGGCCACCGGGTAGCGGTCCGGACACCGGGGTGCACCTGGCCCA | 231 |
| QY | 295 | AlaGlyProProGlyGluArgGlyGly-----LysGlySerLysGlySerGlnGly | 311 |
| Db | 232 | GCGGGTCCGCGCTGGATCTAGAGGTGACCCAGGACCGCTGGTCTCCGGAGCTGCAGGC | 291 |
| QY | 312 | ProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAspPro | 331 |
| Db | 292 | CCACCGGTAGCGTGGCGATCCGGGACACCGGGTGCACTGGCCACCGGGTCCGCCT | 351 |
| QY | 332 | GlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProProGlyPhe | 351 |
| Db | 352 | GGATCTAGAGGTGACCCAGGACCGCGCTGGTCTCCGGGACCTGCAGGCCACCGGGTAGC | 411 |
| QY | 352 | GlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuProGly | 371 |
| Db | 412 | CGTGGGATCCGGGACACCGGGTGCACCTGGCGCCACCGCGGTGGATCTAGAGGT | 471 |
| QY | 372 | LeuProGlyValProGlyMetProGlyProLysGlyPro-----Pro | 385 |
| Db | 472 | GACCCAGGACCGCTGGTCTCGGAGCTCGAGGCCACCGGGTAGCGGTGGCGATCCG | 531 |
| QY | 386 | GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro | 405 |
| Db | 532 | GGACACCGGGTGCACCTGGCCCGAGGGTCCGCGCTGGATCTAGAGGTGACCCAGGACCG | 591 |
| QY | 406 | -----AlaProGluAspAsnGlyCysPro | 413 |
| Db | 592 | CTGTGGTCTCCGGACCTGCAGGCCACCG | 621 |

RESULT 2

```

US-08-655-086-1
:
: Sequence 1, Application US/08655086
: Patent No. 5821089
:
: GENERAL INFORMATION:
: APPLICANT: GRUSKIN, ELLIOTT A.
: APPLICANT: BUECHTER, DOUGLAS
: APPLICANT: ZHANG, GUANGHUI
: APPLICANT: CONNOLLY, KEVIN
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: DILWORTH & BARRESE
: STREET: 333 EARLE OVINGTON BOULEVARD
: CITY: NYONDALE
: STATE: NY
:

```

COUNTRY: US
ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/655,086
FILING DATE: 03-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S.
REGISTRATION NUMBER: 32,063
REFERENCE/DOCKET NUMBER: 203-1632
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-228-8484
TELEFAX: 516-228-8516
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3181 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
US-08-655-086-1

| | | |
|------------------------|----------|------------------|
| Alignment Scores: | | |
| Pred. No.: | 6.82e-17 | 3181 |
| Score: | 452.50 | Matches: 99 |
| Percent Similarity: | 52.11% | Conservative: 12 |
| Best Local Similarity: | 46.48% | Mismatches: 53 |
| Query Match: | 15.24% | Indels: 49 |
| DB: | 1 | Gaps: 8 |

US-09-763-712A-2 (1-547) x US-08-655-086-1 (1-3181)

QY 393 -----AlaValValProLeuAlaLeuGlnAsnGluProThrProAla----- 406
Db 1424 GAACAAGGCCCTGGCTGCCCC--GGATTCCAGGGTCTCCCTGGTCTCT 1480
QY 407 -----ProGluAspAsnGlyCysPro 413
Db 1481 CCAGGTGAAGCAGCAACCTGGTGAACAGGGTGTCTCT 1519
RESULT 3
US-08-642-255-50
; Sequence 50, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642.255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 756 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic"
US-08-642-255-50
Alignment Scores:
Pred. No.: 2,46e-17 Length: 756
Score: 448.50 Matches: 99
Percent Similarity: 52.70% Conservative: 18
Best Local Similarity: 44.59% Mismatches: 85
Query Match: 15.11% Indels: 21
DB: 1 Gaps: 5
US-09-763-712A-2 (1-547) x US-08-642-255-50 (1-756)
QY 197 IleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeuMet-----ArgSerArg 214
Db 56 GTGGCGCTGGACCGGTGCTCCACCGGGTGTCCGGGACCTGCAGGCCCGCCAGGTGCGC 115
QY 215 LeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLysLeuValAspSer 234
Db 116 CTGACCGCTGGTCTCCACCGG-----GTGCTCGGGACCTGCAGGCCCGCCAGGNG 166
QY 235 LysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProGlyProArgGly 254
Db 167 CGCCTGGACCGCTGTGTCACCGG-----GTGCTCCG--GGACCTGCAGGCCCGCCAGGT 219

QY 255 ProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsnLysGlyGlnLys 274
Db 220 GGCCTGGACCGGTGTTCCACCGGGTGTCCGGACCTGCAGGCCCGCCAGGTGCGCCT 279
QY 275 GlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGlyProLysGlyPro 294
Db 280 GGACCGGTGTTCCACCGGGTGTCCGGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCG 339
QY 295 AlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGly 314
Db 340 GCTGTCTCACCGGGTGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCGGTGT 399
QY 315 SerArgGlySerPro-----GlyLysProGlyProGlnGlyProSerGlyAspPro 331
Db 400 CCACCGGTGTTCCGGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCGGTGTCACCG 459
QY 332 GlyProProGlyProGlyGlyGluGlyLeuProGlyProGlnGlyProGlyPhe 351
Db 460 GGTGCTCGGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCGGTGTCACCGGTGCT 519
QY 352 GlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuProGly 371
Db 520 CCGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCGGTGTCACCGGTGTCACCGGTG 579
QY 372 LeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyProSer 391
Db 580 CCGGACCGCCCGCCAGGTGCGCCTGGACCGGTGTCACCGGTGTCACCGGTGTCACCG 639
QY 392 GlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGly 411
Db 640 GGC-----CGCCAGGTGCGCCTGGACCGGTGCT 669
QY 412 CysPro 413
Db 670 CCACCG 675
RESULT 4
US-09-029-348-19
; Sequence 19, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: G087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1608
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-19
Alignment Scores:
Pred. No.: 2,54e-16 Length: 1608
Score: 436.50 Matches: 102
Percent Similarity: 47.06% Conservative: 18
Best Local Similarity: 40.00% Mismatches: 67
Query Match: 14.70% Indels: 70
DB: 4 Gaps: 9
US-09-763-712A-2 (1-547) x US-09-029-348-19 (1-1608)
QY 235 LysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProGlyProArgGly 254
Db 59 CAACATGCCCACTTTACAGAGGAA-CTGTAGAAAGAGGCCCGCCAGGAGTAGAGA 117
QY 255 ProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsnLysGlyGlnLys 274

```

Db 118 CCAGTGGAGAAAGGGTCCACAGGCCCCCAGGAGAGATGGTGAAGATGTCACCA 177
QY 275 GlyGluLysGluProGlyProGlyPro----- 285
Db 178 GGCCCTCCTGGTCCACCTGCTCTCTGCGCCCTCGGTGCGGAACTTTGCTGCT 237
QY 286 -----AlaGlyGluArgGlyProLeGlyProAla 295
Db 238 CAGTATGATGAAAGAGTGGACTTGGCCCTGGACCAATGGCTTAATGGACCTAGA 297
QY 296 GlyProGlyGluArgGlyGlySerLysSerLysSerLysSerLysProLysGlySer 315
Db 298 GGCCCACTGGTCCAGCTGGAGGCCCGCCCTCAAGGTTTCCAAAGGACTGCTGTGAG 357
QY 316 ArgGlySerProGlyLys-----ProGlyProGlnGlyPro----- 327
Db 358 CCGTGGTGAACCTGGTCAACCTGGTCTGCAGGTGCACCTGGTCTCATGGCCCGTGGGT 417
QY 328 -----SerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
Db 418 CCGTGGTGGCAACATGAAACCGTGTGAAACTGGTCTCTGCTGCTGTGCTGCTGCT 477
QY 341 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
Db 478 GGTGCTGTGGCCCAAGAGTCTCTAGTGGCCCAAGGATGCTGTGGCGCATAAAGGAGAG 537
QY 361 ProGlyValProGlyProArgGlyLeuPro-----Gly 371
Db 538 CCGGGTGAAGAGGGGCCAGAGGTCTCTCTGGCTTCAAGGGACACAATGGATGGCAAGT 597
QY 372 LeuProGlyVal-----ProGlyMetProGlyProLys 382
Db 598 CTGCCTGGTATCGTGGTCAACATGCTGATCAAGGTGCTCTGGCTCCGTGGGTCTGCT 657
QY 383 GlyProGlyProGlyProGlyProSerGly----- 392
Db 658 GGTCTAGGGGCCCTGCTGCTCTCTGGCCCTGCTGGAAGATGGTGCACCTGGACAT 717
QY 393 ---AlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGly 411
Db 718 CCGTGGTACGGTGGACCTGCTGG-CATTCAGGCGCTCAGGTCACCAAGGCC----- 770
QY 412 Cys---ProPro-----HisTrpLysAsnPheThrAspLysCys 423
Db 771 TGTGCGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 815

RESULT 5
US-08-159-784-1
; Sequence 1, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4031
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-159-784-1

Alignment Scores: 1.99e-15 Length: 4031
Score: 428.00 Matches: 95
Percent Similarity: 46.41% Conservative: 15
Best Local Similarity: 40.08% Mismatches: 49
Query Match: 14.42% Indels: 78
DB: 1 Gaps: 8

US-09-763-712A-2 (1-547) x US-08-159-784-1 (1-4031)
QY 231 LeuValAspSerLysHisGlyGlnLeuLeuLysAsnPheThrIleLeuGlnGlyProPro 250
Db 976 GTGGTCCAGAGCCCCAACACTCACAACTGTCCTCGGA-----GCACAAGGACCCCG 1026
QY 251 GlyProArgGlyPro----- 255
Db 1027 GGACCTCAGGGGCCACCAGGAGGATGGCACTCCAGGAAGGATGGTGAACCGGGTGAC 1086
QY 256 -----ArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGly 269
Db 1087 CCGTGGTGAAGATGGGAGACCGGGTGACACTGGACCTCAAGGCTTTCAGGAGCCCGGGA 1146
QY 270 AsnLysGlyGlnLysGlyGluLysGlyGluPro-----GlyProProGly 284
Db 1147 GATGTGGCCCTCAAGGCGAGAGGAGATCTCTGTTATGGGCCCGCCGAGGACCTCCAGGG 1206
QY 285 ProAlaGlyGluArgGlyProLe----- 292
Db 1207 CCTCCAGGGCCACCAGGACCTCTCTTCAGACAAGACAAGCTGACCTTCATTGACATGGAG 1266
QY 293 -----GlyPro 294
Db 1267 GGATCGGTTTCAGCGGAGACATAGAGAGCCTTAGAGGCCACAGAGGCTTCCCTGGCCCC 1326
QY 295 AlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGly 308
Db 1327 CCGGGGCCCTCGGTGTCCTCCAGGACTTCTGTGACCCAGGACGCTTGGGATCAATGGT 1386
QY 309 SerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSer 328
Db 1387 TCCTATGCACCA---GGACCTGCAGGCCCTTCCTGCTGTACCTGGGAAGGAGGACCCCG 1443
QY 329 GlyAspProGlyProProGlyProGlyLysGluGlyLeuProGlyProGlnGlyPro 348
Db 1444 GGTTCAGGTCCTCCCGGACCTCCAGGTCT-----CCAGGCAAGAGGCGCCCA 1494
QY 349 ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProAsnG--- 367
Db 1495 CCAGGAGTGGCGGCCGAGAAAGGACGTGTGTGTGATGTGGGATCCAGGACCAAGGG 1554
QY 368 -----GlyLeuProGlyLeuProGlyValProGlyMetPro 379
Db 1555 AGCAAGAGGACCTTGGGGCCCATCGTATGCTGCGCAAGCTGCGCTGGCTGCTGCTGCT 1614
QY 380 GlyProLysGlyProProGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 396

```

Db 1615 GGGCCAGTTGGACCCCGAGGACCTCCAGGGCTCCAGGGCCACCGAGGACCA 1665

RESULT 6

US-09-297-269-39

; Sequence 39, Application US/09297269

; Patent No. 6451557

; GENERAL INFORMATION:

; APPLICANT: VAUGHAN, Paul R.

; APPLICANT: GALANIS, Maria

; APPLICANT: RAMSHAW, John A.M.

; APPLICANT: WERKMEISTER, Jerome A.

; TITLE OF INVENTION: A METHOD FOR PRODUCING, IN YEAST, A HYDROXYLATED TRIPLE

; TITLE OF INVENTION: HELICAL PROTEIN, AND YEAST HOST CELLS USEFUL IN SAID

; FILE OF INVENTION: METHOD (As Amended)

; FILE REFERENCE: Q54094

; CURRENT APPLICATION NUMBER: US/09/297,269

; CURRENT FILING DATE: 1999-04-28

; EARLIER APPLICATION NUMBER: P03310

; EARLIER FILING DATE: 1996-10-29

; EARLIER APPLICATION NUMBER: P04306

; EARLIER FILING DATE: 1996-12-19

; EARLIER APPLICATION NUMBER: PCT/AU97/00721

; EARLIER FILING DATE: 1997-10-29

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 39

; LENGTH: 1572

; TYPE: DNA

; ORGANISM: Collagen Type III Alpha I Chain

US-09-297-269-39

Alignment Scores:

| Pred. No.: | Score: | 1.14e-15 | Length: | 1572 |
|------------------------|--------|---------------|---------|------|
| Percent Similarity: | 42.48% | Matches: | 119 | |
| Best Local Similarity: | 35.10% | Conservative: | 25 | |
| Query Match: | 14.30% | Mismatches: | 108 | |
| DB: | 4 | Gaps: | 15 | |

US-09-763-712A-2 (1-547) x US-09-297-269-39 (1-1572)

QY 224 ValIleMetGluMetLysLeuValaspSerLysHisGlyGlnLeuIleLysAsnPhe 243
 Db 194 GTGAACCTGGAGAGATGGAACCTGGAT-----CAGATGGTC-----TTC 235
 QY 244 ThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySer----- 261
 Db 236 CAGGTGCA-GATGGATCTCTGGTGCAGAGGTGATCGTGGTGAATGGCTCTCTGTGT 294
 QY 262 -----GlnGlyProGlyProGlyProThrGlyAsnLysGlyGlnLys 274
 Db 295 GCCCCTGGGGCTCTGGTTCATCCAGGCCACCTGGTCTCTGGTCCAGCTGGAAGAGCT 354
 QY 275 GlyGluLysGlyGluProGlyPro-----ProGlyProAlaGlyGlu 288
 Db 355 GTGACAGAGAGAGAAGTGGCCCTGCTGGCCCTGCTGGTGCCTGGTCTCTGGTTC 414
 QY 289 ArgGlyProIleGlyProAlaGlyProGlyGluAtgGlyGlyLysGlySerLysGly 308
 Db 415 CGAGGTGCTCTGGTCTCTCAAGGCCACCTGGTGAAGAGGTGAACAGGTGAACCTGGA 474
 QY 309 SerGlnGlyProLysGlySerArg-----GlySerProGlyLysPro 322
 Db 475 GCTGTGGCATCAAGAGCATCGAGGATTCCTGGTGAATCCAGGTGCCCGAGTCTTCCA 534
 QY 323 GlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlyLeu 342
 Db 535 GGCCCTGTGGTGCAGAGGGTGAATCGGCAGTCCAGGACCTGCAGGCCCGCCAGAGGACCT 594
 QY 343 ProGlyProGlnGlyProGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGly 362
 Db 595 GTTGACCCAGTGCACCTCTCTGGCAAGAGATGGAACAGTGGACATCCAGGTCCCATTTGA 654

QY 363 ValProGlyPro-----ArgGlyLeuProGlyLeuProGlyValPro 376
 Db 655 CCACCAGGCGCTCGAGTAACAGAGGTGAAGAGGATCTGAGGCTCCCCAGGCCACCCA 714
 QY 377 GlyMetProGlyProGlyGlyProGlyProGlyProGlyProSer-----Gly 392
 Db 715 GGGCAACAGGCGCTCTGGACCTCTCTGGTGGCCCTGGTGGTGGTGGTGGTGGT 774
 QY 393 AlaValProLeuAlaLeuGlnAsnGlu-----ProThrProAla 406
 Db 775 GCGCTGCCATTCGCTGGGATTCGAGGTGAAGAGGCTGGCGGTTTCCCGCTATTATGGA 834
 QY 407 ProGluAspAsnGlyCysProHisTrpLysAsnPheThrAsp----- 421
 Db 835 CCTGAA-----CCAATGGATTTCAAAAATCAACACCGATGAGATTATCACTTCA 882
 QY 422 ---LysCysTyTyTyPheSerValGluLysGluIlePheGluAspAlaLys----- 437
 Db 883 CTCAAGTCTGTTAATGGACAAATAGAACCTCATTTAGTCTGCTGATGGTTCCTGTAATAAC 942
 QY 438 -----LeuPheCysGluAspLys----- 443
 Db 943 CCGCTAGAACTGCAGAGACCTGAATTCCTGCCATCTCAAGCTCAAGACTCGAGATAC 1002
 QY 444 -----SerSerHisLeuValPheIleAsnThrArg 453
 Db 1003 TGGGTGCAGCCCTAACCAAGGATGCAAAATTTGGATGCTATCAAGGTATTTCTGTAATATGGA 1062
 QY 454 GluGluGlnGlnTrpIleLysLysGlnMetValGly---ArgLysSerHisTrpIleGly 472
 Db 1063 ACTGGGGAACATGATGATGCTGCTGAGAGAAACACGCTTTGGTTGGAGAGTCCATGATGT 1173
 RESULT 7
 US-08-555-669-11
 ; Sequence 11, Application US/08555669
 ; Patent No. 5773248
 ; GENERAL INFORMATION:
 ; APPLICANT: Brewton, Richard G.
 ; APPLICANT: Mayne, Richard
 ; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
 ; NUMBER OF SEQUENCES: 32
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/555,669
 ; FILING DATE: 13-NOV-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Halluin, Albert P.
 ; REGISTRATION NUMBER: 25,227
 ; REFERENCE/DOCKET NUMBER: 8389-030
 ; TELEPHONE: 415-854-3660
 ; TELEFAX: 415-854-3694
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2543 base pairs
 ; TYPE: nucleic acid

```
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..2098
US-08-555-669-11

Alignment Scores:
Pred. No.: 3,85e-15 Length: 2543
Score: 419.00 Matches: 121
Percent Similarity: 35.59% Conservative: 26
Best Local Similarity: 29.30% Mismatches: 91
Query Match: 14.11% Indels: 176
DB: 1 Gaps: 9

US-09-763-712a-2 (1-547) x US-08-555-669-11 (1-2543)

QY 232 ValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProProGly 251
DB 986 CTCGATGGCCAGGAAGGAGAGGCTGGTCCGAAC-----GGTGCTCCGGGA 1030
QY 252 ProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGlyAsnLys 271
DB 1031 GAGAAGGGCCCCAACGGGTGCGGGCTCCCTGGAGAGCGGGGTCCAAAGCGGAGAAG 1090
QY 272 GlyGlnLysGlyGlnLysGlyGlnGlyProProGlyProAlaGlyGlnArgGlyPro 291
DB 1091 GGAGACGGGGCAGAGCTGGGAGCTGGGTGAGGCGCGCCCTCTGGAGAGCCAGCGCTC 1150
QY 292 IleGlyProAlaGlyProProGlyGlnArgGlyGlyLysGlySerLysGlySer----- 309
DB 1151 CTTGGAGATGCTGGCATGCTGGGGAGCGCGGTGAGGCTGGCCACCGGGCTCAGCGGGG 1210
QY 310 -----GlnGlyProLys 313
DB 1211 GCCTCGGGCCCAAGGCCCTCCCGAGGCCCTGGTGTCCGAGGCTTCAGGGCCCAAGAG 1270
QY 314 GlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyPro 333
DB 1271 GGCAGCATGGGAGACCCCGGCTTCAGGCCCCCAGGGCTCCGAGGTGACGTGGCGCAG 1330
QY 334 ProGlyProGlyLysGlnGlyLeuProGlyProGlnGlyProProGlyPheGlnGly 353
DB 1331 CGGGGTCCGGGAGTCCCAAGGCCCTTAAGGAGGACACAGGATTCAGGTTCCGACGTT 1390
QY 354 LeuGlnGlyThrValGlyGlu-----GlnGlyProLys 360
DB 1391 CTTCTGGGATAAAGAGAACTGGGTCCCGGGCTCGTGGACCCCAAGAGAGTCT 1450
QY 360 ----- 360
DB 1451 GGCAGTCGAGGGAGCTGGGGCCCCCAAGGACCCAGGGTCCCAAGCGCACCGCGTGT 1510
QY 361 -----ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly 377
DB 1511 CAGGGTGTCCCGGGCCCCCGGCTCCTGGGCTTCAGAGGCGTCCCGGGTGTCTCTGTC 1570
QY 378 Met----- 378
DB 1571 ATCAGCGGAAGCCGGAGTTCGGGGGAAGGAGGCCAGCAGCAGCGCATCAGGAGCTG 1630
QY 378 ----- 378
DB 1631 TGTGGGGGATGATCAGCGAACAATAATTCACAGTTAGCGCGCACCTAAGGAAGCCTTTG 1690
QY 379 -----ProGlyProLysGlyProProGlyProProGlyProSer 391
DB 1691 GCACCGGGTCCATTGGTCCGGCCCGGTCCAGTGGCCCCCTGGGCCCGCCAGGACCCCA 1750
QY 392 GlyAlaVal-ValProLeuAlaLeuGlnAsnGluPro----- 403
DB 1751 GGCTCATTTGGTCACCTCCGCTCGAGGACCCCGGATACCGCGGTCCACTCGGGAG 1810
```

```
QY 404 -----ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAs 421
DB 1811 CTGGGAGACCCCGGGCCAGAG----- 1832
QY 421 pLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysG1 441
DB 1833 -----GAAACCAGGGTGCAGAGAGAGCAAG----- 1859
QY 441 uAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLy 461
DB 1860 -----GCGCGCAGGAGCAG----- 1874
QY 461 sGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnG1 481
DB 1875 -----GGCTGGACGGGCGCTGAAGGAGACACGAGGGCCCCCAAGGA 1912
QY 481 uTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAs 501
DB 1913 CCCCAGAGC-GTCCCGGCACCAAGCAAGGACGCGGCGGAGCGGTGCTCCCGCGAGCTGG 1971
QY 501 pasnTrpGlyHisGlyHisGlyProGly----- 510
DB 1972 GCCTCCCGGAGATCTCGGGCTTCAGGTGCCATGTGGGGCCAGGGGACACCGCGGGATCTG 2031
QY 511 -----GluAspCysAlaGlyLeuIleTyrAlaGly 520
DB 2032 CGACACCTCAGCTCGCAAGGAGCCGCTGTAGGAGGG 2068

RESULT 8
US-09-073-663-11
; Sequence 11, Application US/09073663
; Patent No. 6127523
; GENERAL INFORMATION:
; APPLICANT: Brewton, Richard G.
; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McGregor & Adler, LLP
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS 8.1
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/09/073,663
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benjamin Aaron Adler, Ph.D., J.D.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5913D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 777-2321
; TELEFAX: (713) 777-6908
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..2098
US-09-073-663-11
```


Alignment Scores:
Pred. No.: 3,85e-15 Length: 2543
Score: 419.00 Matches: 121
Percent Similarity: 35.59% Conservative: 26
Best Local Similarity: 29.30% Mismatches: 91
Query Match: 14.11% Indels: 176
DB: 3 Gaps: 9

US-09-763-712A-2 (1-547) x US-09-073-663-11 (1-2543)

QY 232 ValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProProGly 251
Db 986 CIGATGGCCAGAGGAGGAGGCTGCTCCACAC-----GGTCTCCGGGA 1030

QY 252 ProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGlyAsnLys 271
Db 1031 GAGAGGGGCCCAACGGGCTGCGGGCCCTCCCTGGAGAGCGGGTCCAAAGGCGAGAG 1090

QY 272 GlyGlnLysGlyGluLysGlyGlnProGlyProGlyProAlaGlyGluArgGlyPro 291
Db 1091 GGAGAACGGGGCAGAGCTGGGAGCTGGGTGAGCGCGGCCCTCTGGAGAGCAGCGCTC 1150

QY 292 IleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySer----- 309
Db 1151 CTGGAGATGCTGGCATGCTGGGGAGCGCGGTGAGGCTGGCCACCGGGCTCAGCGGG 1210

QY 310 -----GlnGlyProLys 313
Db 1211 GCCCTCGGCCACAGGCCCTCCGGAGCCCTGGTGTCCGAGGCTTCAGGGCCACAGAG 1270

QY 314 GlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyPro 333
Db 1271 GGCAGCATGGGAGACCCCGGCTTCAGGCGCCCGAGGCTCCGAGGTGACGTGGCGGAC 1330

QY 334 ProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProGlyPheGlnGly 353
Db 1331 CGGGTTCGGGAGGTCGGAAGCCCTTAAGGGAGACAGGGATTCGAGGTTCGCGAGCT 1390

QY 354 LeuGlnGlyThrValGlyGlu----- 360
Db 1391 CTTCTGGGATTAAGAGAGACTGGTCCAGCGGCTGTGTCGAGCCCAAGGAGAGTCT 1450

QY 360 ----- 360
Db 1451 GGCAGTCGAGGGAGCTGGGCCCCAAAGGCACCCAGGGTCCCAAGGCCACAGCGTGT 1510

QY 361 -----ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly 377
Db 1511 CAGGGTGTCCCGGGCCCCCGGCTCCTCTGGGCTGACGGGCTCCCGGGTGTCTCTGC 1570

QY 378 Met----- 378
Db 1571 ATCAGCGGGAGCCGGAGTTCGGGGAGAGGAGCCAGCAGCGCATCAGGGAGCTG 1630

QY 378 ----- 378
Db 1631 TGTGGGGGATGATCAGCAACAAATTGCACAGTTAGCCGCGCACCTAAGGAAGCCTTTG 1690

QY 379 -----ProGlyProLysGlyProProGlyProGlyProGlyProSer 391
Db 1691 GCACCGGGTCCATTGGTCGGCCCGGTCCAGTGGCCCGCTGGGCCCCCAGGACCCCA 1750

QY 392 GlyAlaVal-ProLeuAlaLeuGlnAsnGluPro----- 403
Db 1751 GGCTCCATTGGTCACTCCCTGGGCTGAGAGACCCCGGATACCGCGGTCCCTACTGGGGAG 1810

QY 404 -----ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAs 421
Db 1811 CTGGGAGACCCCGGGCCAGAG----- 1832

QY 421 pLysCysTyrTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysG 441
Db 1833 -----GAAACCAAGGCTGACAGAGAGCAAAAG----- 1859

QY 441 uAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLy 461
Db 1860 -----GGCGCGCAGGAGCAG----- 1874

QY 461 sGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnG 481
Db 1875 -----GGCTGGACGGGCTGAAGGAGCAGGAGGCGGCCCAAGGA 1912

QY 481 uTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAs 501
Db 1913 CCCAAGGC-GTCCCGCGCAGCAGCAGGAGCAGGCGGAGGCTGCTCCCGCGGAGCTGG 1971

QY 501 pAsnTrpGlyHisGlyHisGlyProGly----- 510
Db 1972 GCCTCCCGGAGATCCTGGGCTTCAGGTGCTGCAATTGGGCGCCAGGACACCGGGGATCTG 2031

QY 511 -----GluAspCysAlaGlyLeuIleLysAlaGly 520
Db 2032 CGACACCTCAGCTGCTGCAAGGAGCGCTGTAGGAGGG 2068

RESULT 9
US-08-642-255-48
; Sequence 48, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESS: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic"
US-08-642-255-48

Alignment Scores:
Pred. No.: 1.34e-15 Length: 432
Score: 412.50 Matches: 80
Percent Similarity: 55.70% Conservative: 3
Best Local Similarity: 53.69% Mismatches: 57
Query Match: 13.89% Indels: 9
DB: 1 Gaps: 3

US-09-763-712A-2 (1-547) x US-08-642-255-48 (1-432)

```
QY 248 GlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 267
Db 10 GGGCCAGAGTCCGAGGGCGCGATGTCACAGAGGCCCGGAAAGGTCCGGCTGGACCG 69
QY 268 ThrGlyAsnLysGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGly 287
Db 70 GCTGGTCCACCGGCTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCTGGACCGGCTGT 129
QY 288 GluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLys 307
Db 130 CCACCGGGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCTGGACCGGCTGCACCG 189
QY 308 GlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyPro 327
Db 190 GGTGCTCCGGGACCT-----GCAGGCCCGCCAGGTGCGCTGGACCGGCTGTCAC 240
QY 328 SerGlyAspProGlyProProGlyProProGlyLysGluGlyLysProGlyProGlnGly 347
Db 241 CCGGGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCTGGACCGGCT 342
QY 368 GlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyPro 387
Db 343 GGTCCACCGGGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCTGGACCGGCTGTCAC 402
QY 388 ProGlyProSerGlyAlaValPro 396
Db 403 AAGGAGCTCAGGTCCTCCGGCAGGTCCG 429
```

RESULT 10

US-08-159-784-4

```
; Sequence 4, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55sx
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3394
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-159-784-4
Alignment Scores:
Pred. No.: 1.87e-14 Length: 3394
Score: 409.00 Matches: 86
Percent Similarity: 48.98% Conservative: 10
Best Local Similarity: 43.88% Mismatches: 42
Query Match: 13.78% Indels: 58
DB: 1 Gaps: 7
US-09-763-712A-2 (1-547) x US-08-159-784-4 (1-3394)
QY 246 LeuGlnGlyProGlyProArgGlyProAlaGlyAspArgGlySer----- 261
Db 553 ATGCCCGGGCCCCCAGGACCTCCAGGGCCCCCAGGCCCTCCAGGACTCTGTTTACGAC 612
QY 262 -----GlnGlyProProGlyProThrGlyAsnLysGly 272
Db 613 AGCAATGTGTTTGTGCTGAGTCCAGCGCCCGGCGCTCCAGGATTGCCAGGAATCAG--- 669
QY 273 GlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGlyProIle 292
Db 670 -----GGCCCTCCAGGACCAAGGGCCCCAAGAGGAAGTG 705
QY 293 GlyProAlaGlyProGlyGlu----- 300
Db 706 GGGCCCCCGGACCCAGGCGAGTTTCCGTTTGACTTTCTCAGAGGAGGCTGAATG 765
QY 301 ArgGlyGlyLysGlySerLysGlnGlyProLysGlySerArgGlySer----- 318
Db 766 AAGGGGAGAGGAGACCGAGGTGATGCAGGACAGAAAGGGGAGGCGGAGCGGG 825
QY 319 -----ProGlyLysProGlyProGlnGlyProSerGly 329
Db 826 GGGCGGGTTCCTTTCGCTGCTCAGCGTGCCTGGGGCCCCCGGCGCCCGCCAGCGCTGGC 885
QY 330 AspProGlyProGlyProGlyLysGlu-----GlyLeuProGly 344
Db 886 TACCCTGGGATTCAGGTCCCAAGGGAGAGCATCCGGGGCCAGCGCCGACCTGGA 945
QY 345 ProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValPro 364
Db 946 CCTCAGGGACCCCGCGCATC---GGCTACGAGGGGCGCAGGGCCCTCCCGGCCCCCA 1002
QY 365 GlyProArgGlyLeuProGlyLeuProGly-----ValProGly 377
Db 1003 GGGCCCCCAGGGCCCCCTTCATTTCTGCGCCCTCAGGACAGACTATCAGCGTTCCGGG 1062
QY 378 MetProGlyProLysGlyProGlyProGlyProGlyProSerGlyAla 393
Db 1063 CCTCGGGCCCCCTTGGGGCCCCCTGGGCCCTTGAACCATGGCGCC 1110
```

RESULT 11

```
US-07-621-091G-1
; Sequence 1, Application US/07621091G
; Patent No. 5424408
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen T., Morrison, Karen E., Hudson, Billy
; APPLICANT: G.
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen
; TITLE OF INVENTION: Polynucleotides
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University, Office of Cooperative Research
; STREET: 246 Church Street
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800K storage
```

; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: Macintosh OS7.0
 ; SOFTWARE: Microsoft Word 5.1a
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/621,091G
 ; FILING DATE: 11/30/90
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA: NO. 5424408 applicable
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Barth, Richard S.
 ; REGISTRATION NUMBER: 28180
 ; REFERENCE/DOCKET NUMBER: 900983/RB
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 972-1400
 ; TELEFAX: (212) 370-1622
 ; TELEX: 236268
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1416 base pairs
 ; TYPE: Nucleic acid
 ; STRANDEDNESS: Double
 ; TOPOLOGY: Linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Calf
 ; STRAIN: Unknown
 ; INDIVIDUAL ISOLATE: Unknown
 ; DEVELOPMENTAL STAGE: Unknown
 ; CELL TYPE: Whole kidney
 ; IMMEDIATE SOURCE:
 ; LIBRARY: Bovine lens cDNA
 ; CLONE: KMC15
 ; POSITION IN GENOME: NO. 5424408 known
 ; CHROMOSOME/SEGMENT: NO. 5424408 known
 ; PUBLICATION INFORMATION: NO. 5424408e
 ; US-07-621-091G-1

Alignment Scores:
 Pred. No.: 1416
 Score: 402.50
 Percent Similarity: 46.62%
 Best Local Similarity: 38.72%
 Query Match: 13.56%
 DB: 1
 Gaps: 8

US-09-763-712A-2 (1-547) x US-07-621-091G-1 (1-1416)

QY 247 GlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGly 266
 Db 112 AAAGGAGATCGAGGTCCACCTGCTCCAGAGAAACCCAGGCATGCTGCTCTCTGGA 171
 QY 267 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyProGly 281
 Db 172 CCTCCAGGAGGAGTCTGTAGAGGCATATAAAGGAGACAGGGGTTGATGGAGAGCCTGGC 231
 QY 282 ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArg 301
 Db 232 CAAGAGGTCCACCTGAGCTATAGAGACATGGGGTCCAGGTCATCCGGAGCAGCA 291
 QY 302 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 321
 Db 292 GGTGTCCCGGTCCAGCCAGGCGCAGAGGTGATCCTGGATTCTATGGATTCCAGGCATG 351
 QY 322 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyLysGlyGlu 341
 Db 352 AAAGGAGAAAGGGTAATTTCAGGATTTCCAGGACACCTGGACCTCCAGGCAAGTGA 411
 QY 342 -----LeuPro 343
 Db 412 CCAAAAGGACCACCTGGAGTACGTGGAGAGCCTGGCACAGTGAAGATCATCTCCCTTCCA 471

QY 344 GlyProGlnGlyProGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 363
 Db 472 GGAAGCCCGAGCCACCTGTTGCTGAGTGGAGAACCCAGGATGCAAGGAGAAACCCGGGCC 531
 QY 364 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 383
 Db 532 CCAGGACCACCGAGGATCCAGGACCTGTGGGCCAAAAGGTAAACCCAGGGGAGGATGGT 591
 QY 384 ProGlyProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 403
 Db 592 CCACCAGGAACCTCTGACCACTGGAGAAAAGGCAACAAAGGTTGTTAAAGGAGAGCAA 651
 QY 404 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 419
 Db 652 GGACCA---CCTGGATCCGATCCGCTGCCAGGCTTGAAGGGGAAACCTGAGACACACTGG 707
 QY 420 ThrAspLysCysTyrTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 439
 Db 708 ACC---ACCTGCAGCAGGGGCGATGATGAGGGGCTTTGTCTTTACCCGGCACAG----- 758
 QY 440 CysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 459
 Db 759 ---CCAGACCACAGCAATTCCTCTGTCAGAGAGGAGCAGACCCGCTCTATAGTGGTT 815
 QY 460 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 477
 Db 816 TTCTCTCTCTTTGTACAAAGAAATGAACAAGCCCATGGACAGGA----- 860
 QY 478 ArgGluAsnGluTrpLys 483
 Db 861 CCTGGGACACTTGGCAG 878
 RESULT 12
 ; US-08-399-889-1
 ; Sequence 1, Application US/08399889B
 ; Patent No. 5973120
 ; GENERAL INFORMATION:
 ; APPLICANT: Reeder, Stephen T
 ; APPLICANT: Morrison, Karen E
 ; APPLICANT: Hudson, Billy G
 ; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
 ; FILE REFERENCE: 951263A
 ; CURRENT APPLICATION NUMBER: US/08/399,889B
 ; CURRENT FILING DATE: 1995-03-07
 ; EARLIER APPLICATION NUMBER: 07/621091
 ; EARLIER FILING DATE: 1990-11-30
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1416
 ; TYPE: DNA
 ; ORGANISM: Calf
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1416)
 ; US-08-399-889-1
 Alignment Scores:
 Pred. No.: 1,69e-14
 Score: 402.50
 Percent Similarity: 46.62%
 Best Local Similarity: 38.72%
 Query Match: 13.56%
 DB: 2
 Gaps: 8
 US-09-763-712A-2 (1-547) x US-08-399-889-1 (1-1416)
 QY 247 GlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGly 266
 Db 112 AAAGGAGATCGAGGTCCACCTGCTCCAGAGAAACCCAGGCATGCTGCTCTCTGGA 171
 QY 267 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyProGly 281
 Db 172 CCTCCAGGAGGAGTCTGTAGAGGCATATAAAGGAGACAGGGGTTGATGGAGAGCCTGGC 231

Db 172 CCTCCAGGAGTCTCTAGAGGCATATAAGAGGAGCAAGGGGTGTATGGAGAGCCTGGC 231
 QY 282 ProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArg 301
 Db 232 CAAAGAGTCCACCTGGAGCTATAGGAGATGGGTACCAGTCCATCCGGAGCACCA 291
 QY 302 GlyGlyLysGlySerLysGlyGlnGlyProLysGlySerArgGlySerProGlyLys 321
 Db 292 GGTGTCCCGGTGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 351
 QY 322 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLys 341
 Db 352 AAAGGAAGAGGGTAATTCAGATTTCAGAGGACCTGGACCTCCAGGCAAGAGTGA 411
 QY 342 -----LeuPro 343
 Db 412 CCAAAAGGACCACTGGAGTACGTGGAGAGCCTGGACAGTGAAGATCATCTCCCTTCCA 471
 QY 344 GlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 363
 Db 472 GGAAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 531
 QY 364 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys 383
 Db 532 CCAGGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 591
 QY 384 ProProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 403
 Db 592 CCACCAAGGAAGTCCCTGGACCACTGGAGAAAGGAGGAGGAGGAGGAGGAGGAG 651
 QY 404 ThrProAlaProGluAsnGlyCysPro-----ProHisTrpLysAsnPhe 419
 Db 652 GGACCA---CCTGGATCCGATGGCTGCCAGGCTTGAAGGGGAAACCTGAGACACTGG 707
 QY 420 ThrAspLysCysTrpTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 439
 Db 708 ACC---ACCTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 758
 QY 440 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 459
 Db 759 ---CCAGACCAAGCAATTCCTCTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 815
 QY 460 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 477
 Db 816 TTCTCTCTCTTTGTACAAGAAATGAACAAGCCATGGACAGCA-----860
 QY 478 ArgGluAsnGluTrpLys 483
 Db 861 CCTGGGAACACTGGCAG 878

RESULT 13
 US-09-167-364-1
 ; Sequence 1, Application US/09167364
 ; Patent No. 6007980
 ; GENERAL INFORMATION:
 ; APPLICANT: Reeder, Stephen T
 ; APPLICANT: Morrison, Karen E
 ; APPLICANT: Hudson, Billy G
 ; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
 ; FILE REFERENCE: 951263B
 ; CURRENT APPLICATION NUMBER: US/09/167,364
 ; CURRENT FILING DATE: 1998-10-07
 ; EARLIER APPLICATION NUMBER: 08/399889
 ; EARLIER FILING DATE: 1995-03-07
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 1416
 ; TYPE: DNA
 ; ORGANISM: Calf
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1416)

US-09-167-364-1
 Alignment Scores:
 Pred. No.: 1-69e-14 Length: 1416
 Score: 402.50 Matches: 103
 Percent Similarity: 46.62% Conservative: 21
 Best Local Similarity: 38.72% Mismatches: 103
 Query Match: 13.56% Indels: 40
 DB: 3 Gaps: 8

US-09-763-712A-2 (1-547) x US-09-167-364-1 (1-1416)
 QY 247 GlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGly 266
 Db 112 AAAGAGATCGAGGTCCACCTGGCTCCAGAGAAACCCAGGATGCTGCTGCTCTCTGGA 171
 QY 267 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 281
 Db 172 CCTCCAGGAGGAGTCCCTAGAGGCATATAAGAGGAGCAAGGGGTGTATGGAGAGCCTGGC 231
 QY 282 ProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArg 301
 Db 232 CAAAGAGTCCACCTGGAGCTATAGGAGATGGGTCCACAGTCCATCCGGAGCACCA 291
 QY 302 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 321
 Db 292 GGTGTCCCGGTGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 351
 QY 322 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGly 341
 Db 352 AAAGGAAGAGGGTAATTCAGATTTCAGGACCTGGACCTCCAGGCAAGAGTGA 411
 QY 342 -----LeuPro 343
 Db 412 CCAAAAGGACCACTGGAGTACGTGGAGAGCCTGGACAGTGAAGATCATCTCCCTTCCA 471
 QY 344 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 363
 Db 472 GGAAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 531
 QY 364 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys 383
 Db 532 CCAGGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 591
 QY 384 ProProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 403
 Db 592 CCACCAAGGAAGTCCCTGGACCACTGGAGAAAGGAGGAGGAGGAGGAGGAGGAGGAG 651
 QY 404 ThrProAlaProGluAsnGlyCysPro-----ProHisTrpLysAsnPhe 419
 Db 652 GGACCA---CCTGGATCCGATGGCTGCCAGGCTTGAAGGGGAAACCTGAGACACTGG 707
 QY 420 ThrAspLysCysTrpTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 439
 Db 708 ACC---ACCTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 758
 QY 440 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 459
 Db 759 ---CCAGACCAAGCAATTCCTCTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 815
 QY 460 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 477
 Db 816 TTCTCTCTCTTTGTACAAGAAATGAACAAGCCATGGACAGCA-----860
 QY 478 ArgGluAsnGluTrpLys 483
 Db 861 CCTGGGAACACTGGCAG 878

RESULT 14
 US-09-439-897-1
 ; Sequence 1, Application US/09439897
 ; Patent No. 6277558
 ; GENERAL INFORMATION:

APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 95-1263-C
CURRENT APPLICATION NUMBER: US/09/439,897
CURRENT FILING DATE: 1999-11-12

NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 1416

TYPE: DNA

ORGANISM: Bos taurus

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(1416)

US-09-439-897-1

Alignment Scores:

Pred. No.: 1.69e-14 Length: 1416

Score: 402.50 Matches: 103

Percent Similarity: 46.62% Conservative: 21

Best Local Similarity: 38.72% Mismatches: 103

Query Match: 13.56% Indels: 40

DB: Gaps: 8

US-09-763-712A-2 (1-547) x US-09-439-897-1 (1-1416)

QY 247 GlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGly 266

Db 112 AAAGGATCGAGTCCACTGCTCCAGAGGAACCCAGGATGCTGCTCTCTGGA 171

QY 267 ProThrGlyAsn-----LysGlnGlnLysGlyGluLysGlyProGly 281

Db 172 CCTCAGGAGTCTCTAGAACGATATAAAGAGACAAAGGGTTGATGGAGAGCCTGTC 231

QY 282 ProGlyProAlaGlyGluArgGlyProGlyProAlaGlyProGlyGluArg 301

Db 232 CAAAGAGTCCACTGGAGCTATAGGACATCCAGGATCATCCGGAGCACCA 291

QY 302 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 321

Db 292 GGTGTCCCGTCCAGCCAGGGCCAGAGGTGATCTGGATTCTATGATTTCCAGCATG 351

QY 322 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyLysGluGly 341

Db 352 AAAGGAAGAGGAATTCAGATTTCAGGACCACTGGACCTCCAGGGCAAAAGTGA 411

QY 342 -----LeuPro 343

Db 412 CCAAAGGACCACTGGAGTACGTGGAGAGCCTGGCACAGTGAAGATCATCTCCCTTCA 471

QY 344 GlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 363

Db 472 GGAAGCCAGCCACCTGTTGAGTGGAGAACAGGATGCAAGGAACCCGGGCC 531

QY 364 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 383

Db 532 CCAGGACCAACAGGAGATCCAGGACCTGTGGGCCCAAAAGGTAACACAGGGAGGATGT 591

QY 384 ProGlyProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 403

Db 592 CCACCAAGGACCTCTGGACCACTGGAGAAAAGGCAACAAAGTTGTAAAGAGAGACAA 651

QY 404 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 419

Db 652 GGACCA--CCTGGATCCGATGCGCTTCAAGGGGAAACC-TGGAGACACTGG 707

QY 420 ThrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 439

Db 708 ACC---ACCTGCAGCAGGGGAGTATGATGAGGGCTTTGTCTTTACCCGCACAG----- 758

QY 440 CysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 459

Db 759 ----CCAGACCAAGCAATTCCTCTCTGCTCCGAAAGGACAGAGCCGCTCTATAGTGGT 815

QY 460 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 477

Db 816 TTCTCTCTCTCTTTGTACAAAGAAATGAACAAGCCCATGGACAGCA----- 860

QY 478 ArgGluAsnGluTrpLys 483

Db 861 CCTGGGAACACTTGGCAG 878

RESULT 15

US-09-484-970B-4

Sequence 4, Application US/09484970B

Patent No. 6426186

GENERAL INFORMATION:

APPLICANT: Jones, Karen A.

APPLICANT: Volkmuth, Wayne

TITLE OF INVENTION: BONE REMODELING GENES

FILE REFERENCE: PB-0014 US

CURRENT APPLICATION NUMBER: US/09/484,970B

CURRENT FILING DATE: 2000-01-18

NUMBER OF SEQ ID NOS: 172

SOFTWARE: PERL Program

SEQ ID NO 4

LENGTH: 4359

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

OTHER INFORMATION: Incyte ID No. 6426186 331365 (1553795CB1)

US-09-484-970B-4

Alignment Scores:

Pred. No.: 7.21e-14 Length: 4359

Score: 400.50 Matches: 90

Percent Similarity: 54.89% Conservative: 11

Best Local Similarity: 48.91% Mismatches: 64

Query Match: 13.49% Indels: 19

DB: Gaps: 7

US-09-763-712A-2 (1-547) x US-09-484-970B-4 (1-4359)

QY 246 LeuGlnGlyProGlyProArgGlyProArgGlyAspArg----- 259

Db 448 CTGCCAGGACCCAGGCCCCAGGTCGCAAGAGGCTCCGAGGAATCCAGGCTTCGCA 507

QY 260 GlySerGlnGlyProGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGlu 279

Db 508 GGAGCTGATGGAGACCCAGGGCTCCAGGAGACGCGAGGTCGTGAAGGGTTC 567

QY 280 ProGlyProPro-----GlyProAlaGlyGluArgGlyProIleGlyProAlaGly 296

Db 568 CCAGGACCCAGGGTTCATAGGACCCCGAGGATCCAAAGGTCAGTGGGCTCCCTGGC 627

QY 297 ProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArg 316

Db 628 CCAGATGGATCCCAAGTCCCATCGGCTGCCAGGCGCAGATGGGCCCTCGGGGAAAG 687

QY 317 GlySerProGlyLysPro---GlyProGln---GlyProSerGlyAspProGlyProPro 334

Db 688 GGCTCTCTGGAGAGTCTCTGGAGATCCAGCCCGGCGCACGGGAGATGCTGTGTGCT 747

QY 335 GlyProGlyLysGluGlyProGlyProGlnGlyProGlyPheGlnGlyLeu 354

Db 748 GGACAGCTGGGCTTAAGGCTTCCCGAGACAGAGGCCCCCTGGATTCAGAGGAGC 807

QY 355 GlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuProGlyProGly 374

Db 808 CAAGGATG-----CCTGGGATGTCAGGCTGAAGGGCCAGCCAGGCTCCAGGA 858

QY 375 ValProGlyMetProGlyProLysGlyProGlyProProGlyPro-----ProGlyProSer 391

Db 859 CCTTCCGGCCAGCCAGGCTGTATGGGCTCCAGGACTGCATGATGATTCAGGAGCTCCT 918

```

.QY 392 GlyAlaValProLeuAlaLeuGlnAsnGluPro-----ThrProAlaProGluAsp 409
Db 919 GGCCAAAGAGGGCCCTTGGGGCTGCCAGGAATCCACAGGCCGTGAAGGTCTGCTGTGTGAT 978
.QY 410 AsnGlyCysPro 413
Db 979 AGAGGGGACCCT 990

```

Search completed: March 21, 2003, 12:34:04
Job time : 73.1137 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:23:12 ; Search time 265.779 Seconds
(without alignments)
1596.010 Million cell updates/sec

Title: US-09-763-712A-2
Perfect score: 2969
Sequence: 1 MYSHNVYIMNLLNLTVQV.....EDVNNFCEKDRVTLSSAL 547

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09763712/runat_14032003.100950.18119/app_query.fasta_1.1877
-DB=Published Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09763712@cgn.1.123 @runat_14032003.100950.18119
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:
1: /cgn2.6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2.6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2.6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2.6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2.6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2.6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2.6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2.6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2.6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2.6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2.6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2.6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2.6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2.6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|--------------------|
| 1 | 2936 | 98.9 | 2930 | 10 | US-09-745-763-198 |
| 2 | 474.5 | 16.0 | 1410 | 10 | US-09-954-456-2274 |
| 3 | 455 | 15.3 | 1133 | 9 | US-09-924-340-57 |
| 4 | 455 | 15.3 | 1133 | 9 | US-09-992-600A-57 |
| | | | | | Sequence 198, App |
| | | | | | Sequence 2274, App |
| | | | | | Sequence 57, Appl |

| | | | | | | |
|----|-------|------|------|----|--------------------|--------------------|
| 5 | 453 | 15.3 | 6158 | 10 | US-09-919-497-6 | Sequence 6, Appl |
| 6 | 453 | 15.3 | 6158 | 10 | US-09-954-456-762 | Sequence 762, App |
| 7 | 452.5 | 15.2 | 6728 | 10 | US-09-954-456-782 | Sequence 782, App |
| 8 | 452.5 | 15.2 | 6728 | 10 | US-09-880-107-3946 | Sequence 3946, App |
| 9 | 444 | 15.0 | 5416 | 10 | US-09-954-456-786 | Sequence 786, App |
| 10 | 444 | 15.0 | 5416 | 10 | US-09-880-107-2094 | Sequence 2094, App |
| 11 | 440.5 | 14.8 | 4149 | 12 | US-10-044-090-104 | Sequence 104, App |
| 12 | 440.5 | 14.8 | 5086 | 10 | US-09-880-107-3947 | Sequence 3947, App |
| 13 | 440.5 | 14.8 | 5145 | 9 | US-09-925-299-206 | Sequence 206, App |
| 14 | 440.5 | 14.8 | 5145 | 10 | US-09-925-299-206 | Sequence 206, App |
| 15 | 440.5 | 14.8 | 5432 | 12 | US-10-044-090-22 | Sequence 22, Appl |
| 16 | 438 | 14.8 | 2235 | 10 | US-09-962-436-308 | Sequence 308, App |
| 17 | 435 | 14.7 | 3690 | 12 | US-10-044-090-448 | Sequence 448, App |
| 18 | 433 | 14.6 | 2192 | 10 | US-09-925-301-42 | Sequence 42, Appl |
| 19 | 432.5 | 14.6 | 1486 | 10 | US-09-925-302-247 | Sequence 247, App |
| 20 | 431.5 | 14.5 | 3226 | 10 | US-09-954-456-725 | Sequence 725, App |
| 21 | 426.5 | 14.4 | 4908 | 9 | US-10-001-887-33 | Sequence 33, Appl |
| 22 | 425.5 | 14.3 | 2520 | 10 | US-09-880-107-3685 | Sequence 3685, App |
| 23 | 423 | 14.2 | 481 | 10 | US-09-815-343-1516 | Sequence 1516, App |
| 24 | 418.5 | 14.1 | 511 | 9 | US-09-736-457-1138 | Sequence 1138, App |
| 25 | 418.5 | 14.1 | 511 | 9 | US-09-902-941-1138 | Sequence 1138, App |
| 26 | 418.5 | 14.1 | 511 | 9 | US-09-849-626-1138 | Sequence 1138, App |
| 27 | 415.5 | 14.0 | 2823 | 10 | US-09-919-497-7 | Sequence 7, Appl |
| 28 | 414.5 | 14.0 | 2542 | 9 | US-09-954-531-961 | Sequence 961, App |
| 29 | 414.5 | 14.0 | 2542 | 10 | US-09-964-824A-255 | Sequence 255, App |
| 30 | 409 | 13.8 | 3394 | 10 | US-09-880-107-2178 | Sequence 2178, App |
| 31 | 407 | 13.7 | 477 | 10 | US-09-815-343-1517 | Sequence 1517, App |
| 32 | 397.5 | 13.4 | 2691 | 10 | US-09-925-302-64 | Sequence 64, Appl |
| 33 | 394.5 | 13.3 | 1797 | 9 | US-09-978-295A-613 | Sequence 613, App |
| 34 | 394.5 | 13.3 | 1797 | 9 | US-09-978-697-613 | Sequence 613, App |
| 35 | 394.5 | 13.3 | 1797 | 9 | US-09-978-192A-613 | Sequence 613, App |
| 36 | 394.5 | 13.3 | 1797 | 9 | US-09-999-832A-613 | Sequence 613, App |
| 37 | 394.5 | 13.3 | 1797 | 9 | US-09-978-189-613 | Sequence 613, App |
| 38 | 394.5 | 13.3 | 1797 | 9 | US-10-174-590-331 | Sequence 331, App |
| 39 | 394.5 | 13.3 | 1797 | 9 | US-10-176-758-331 | Sequence 331, App |
| 40 | 394.5 | 13.3 | 1797 | 9 | US-10-175-737-331 | Sequence 331, App |
| 41 | 394.5 | 13.3 | 1797 | 9 | US-10-173-706-331 | Sequence 331, App |
| 42 | 394.5 | 13.3 | 1797 | 9 | US-10-175-738-331 | Sequence 331, App |
| 43 | 394.5 | 13.3 | 1797 | 9 | US-10-175-752-331 | Sequence 331, App |
| 44 | 394.5 | 13.3 | 1797 | 9 | US-10-176-482-331 | Sequence 331, App |
| 45 | 394.5 | 13.3 | 1797 | 9 | US-10-176-757-331 | Sequence 331, App |

ALIGNMENTS

RESULT 1
US-09-745-763-198
; Sequence 198 Application US/09745763
; Patent No. US20020065394A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; Lavallie, Edward R.
; Collins-Racie, Lisa A.
; Evans, Cheryl
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; ENCODING THEM
; NUMBER OF SEQUENCES: 219
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/745,763
FILING DATE: 18-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 198:
SEQUENCE CHARACTERISTICS:
LENGTH: 2930 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-09-745-763-198

Alignment Scores:

Pred. No.: 9,15e-151 Length: 2930
Score: 2936.00 Matches: 545
Percent Similarity: 99.63% Conservative: 0
Best Local Similarity: 99.63% Mismatches: 2
Query Match: 98.89% Indels: 1
DB: 10 Gaps: 0

US-09-763-712A-2 (1-547) x US-09-745-763-198 (1-2930)

```
QY 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnLeuAsnLeuThrGlnValGln 20
DB 652 ATGTATCTCATATATGTGTCATCATGAATC-AACAACCTGAACCTGACCCAGGTGCAG 710
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
DB 711 CAGAGAACTCTATCAGAACTGCGAGCGTCTGTGGATGACACAGCCAGGCTATCCAG 770
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
DB 771 CGAATCAAGAAGACGATTTCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGGACACG 830
QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAlaLeu 80
DB 831 GATTTGGCTGAAGAGAGAAAGTGAGAGCTTGCAGACCTGGCTGCCAACAACTCTGCGTTG 890
QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
DB 891 GCCAAAGCCCAACACACACCTGGAGGATATGAACAGCCAGCTCAACTCATTCACAGGT 950
QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
DB 951 CAGATGGAGAATCATCACCACTATCTCTCAAGCCACAGCAGCAACCTGAAAGACCTGCAG 1010
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
DB 1011 GACTTACAAAGATGCAGAGATAGAACGCGCATCAAGTTCAACCACTGGAGGAACGC 1070
QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerThrAlaHis 160
DB 1071 TTCCAGCTCTTTCAGACGGATATTGTGAACATCATTTAGCAATATCATGTTACAGCCAC 1130
QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
DB 1131 CACCTCGGACGCGTGACCAAGCAATCTAAATGAAGTCAGGACCACTTGACAGATACCCCT 1190
QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnThrLeuAlaAsnIleArgLeuAsp 200
DB 1191 ACCAAACACACAGATGATCTGACCTCTTGAATAATATACCTGGCCCAACATCCGTTTGGAT 1250
QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
DB 1251 TCTGTTTCTCTCAGGATGCAACAAGATTTGATGAGGTCAGGCTAGACACTGAAGTAGCC 1310
```

```
QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
DB 1311 AACTTATCAGTGATTATGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAGCTCATC 1370
QY 241 LysAsnPheThrIleLeuGlnGlyProGlyProGlyProArgGlyProArgGlyAspArgGly 260
DB 1371 AAGAATTTTACAATACTACAAGGTCCACGGGGCCCGGGGGTCCAAAGAGGTGACAGAGGA 1430
QY 261 SerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro 280
DB 1431 TCCAGGAGCCCTCTGGCCCACTGGCAACAGGACAGAAAGAGAGAGAGGGAGCCCT 1490
QY 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProAlaGlyProGlyGlu 300
DB 1491 GGACACCTGGCCCTCGGGTGAGAGAGGCCCAATTTGGACCACTGGTCCCCCGGAGAG 1550
QY 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProGlySerGlySerArgGlySerProGly 320
DB 1551 CGTGGCGCAAAAGGATCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTCCCTGGG 1610
QY 321 LysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 340
DB 1611 AAGCCCGGCCCTCAGGGCCCCAGTGGGAGCCCGAGGCCCGGGGCCCGCCAGGCAAGAG 1670
QY 341 GlyLeuProGlyProGlnGlyProGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
DB 1671 GGACTCCCGCGCCCTCAGGGCCCTCTGGCTTCCAGGGACTTCCAGGGCAGCCGTTGGGAG 1730
QY 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
DB 1731 CCTGGGCTGCCCTGGACTCGGGGACTGCCAGGCTTCCCTGGGGTACAGGCATGCCAGGC 1790
QY 381 ProLysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGln 400
DB 1791 CCCAAGGGCCCCCGGGCCCTCTGCCCCATCAGGAGCGGTGTGCCCCCTGGCCCCCTGCAG 1850
QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
DB 1851 AATGAGCCCAACCCCGGCGCAGGAGGACAATAGCTGCCGCCCTCCTCCTGGAAGAACTTCA 1910
QY 421 AspLysCysTyrTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
DB 1911 GACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTGGAGGATCAAAAGCTTTTCTGT 1970
QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLys 460
DB 1971 GAAGACAAGCTTTCACATCTTGTTTTTCATAAACACACTAGAGAGGACAGCAATGATAAAA 2030
QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
DB 2031 AAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAATAAT 2090
QY 481 GluTrpLysTrpLeuAspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnPro 500
DB 2091 GAATGAAGTGGCTGTGATGGGACATCTCCAGACTACAAAAATGGAAAGCTGGACAGCCG 2150
QY 501 AspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleValAlaGly 520
DB 2151 GATACTGGGGTTCATGGCCATGGCCAGGAGAGAGACTGTGCTGGTGTGATTTATGCTGGG 2210
QY 521 GlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlu 540
DB 2211 CAGTGGAAAGCATTTCCAATGTGAAGACGTCAATAACTTTTCATTTTGCAGAAAAAGACAGGAG 2270
QY 541 ThrValLeuSerSerAlaLeu 547
DB 2271 ACAGTACTGTCTCATTCATTA 2291
```

RESULT 2

US-09-745-456-2274
; Sequence 2274, Application US/09954456
; Patent No. US20020115057A1

| | | | |
|--|----------|--|------|
| Alignment Scores: | | | |
| Pred. No.: | 4,31e-18 | Length: | 1410 |
| Score: | 474.50 | Matches: | 111 |
| Percent Similarity: | 49.67% | Conservative: | 41 |
| Best Local Similarity: | 36.27% | Mismatches: | 132 |
| Query Match: | 15.98% | Indels: | 23 |
| DB: | 10 | Gaps: | 6 |
| US-09-763-712A-2 (1-547) x US-09-954-456-2274 (1-1410) | | | |
| QY | 246 | LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro | 265 |
| DB | 400 | ATGCTGGACAAGCTGGCCAGTTGGGCCCAAGGGGACAATGGCTCTGTGGAGAACCT | 459 |
| QY | 266 | GlyProThrGlyAsnLysGlyLynLysGlyGluLysGlyGluProGlyProGlyPro | 285 |
| DB | 460 | GGACCAAGGGAGACACTGGCCCAAGTGACCTCCAGGACCTCCCGTGTGCTGTGTC | 519 |
| QY | 286 | AlaGlyGluArgGlyProLleGlyProAlaGlyProGlyGluArgGlyGlyLysGly | 305 |
| DB | 520 | GCTGGAAAGAAGTCCCTGGGGAAGCAGGGGAACATAGGACCTCAGGGCAAGCCAGGC | 579 |
| QY | 306 | SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln | 325 |
| DB | 580 | CCAAAAGGAAGCTGGGCCAAAGGAGAAGTAGGTGCCCGACAGCATGAGGGCTCGCA | 639 |
| QY | 326 | GlyProSerGlyAspProGlyProGlyPro-----ProGlyLysGluGlyLeu | 342 |
| DB | 640 | GGGCAAGAGCCCTCGAGGCCCTTAGGGAGCAGGAGTGTCCCTGGTGGCGTGGAGTC | 699 |
| QY | 343 | ProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGly | 362 |
| DB | 700 | CCTGGAAACGACGGGCAGCGGTCTGCTGGAGCCATGGTCCCGAGGAGTCCAGGT | 759 |
| QY | 363 | ValProGlyProArgGlyLeuProGlyValProGlyMetProGlyProLys | 382 |
| DB | 760 | GCCAGGGGACCCCGGAGTTGAAGGGGACAAAGGCATTCTCTGACAGCAAGAGCAAG | 819 |

[illegible]

```
Alignment Scores:
Pred. No.: 3,99e-17 Length: 1133
Score: 455.00 Matches: 98
Percent Similarity: 47.46% Conservative: 14
Best Local Similarity: 41.53% Mismatches: 50
Query Match: 15.33% Indels: 74
DB: 9 Gaps: 7

US-09-763-712A-2 (1-547) x US-09-924-340-57 (1-1133)
QY 246 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 265
DB 140 CTCCAGGCCCCCAGGCCCCAGGCGTCCAGAGGAGCGAGGACCCCAAGGTAACCTCC 199
QY 266 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 285
DB 200 GGTGAGAAGGGCCACCGAGGATTCAGGCGCCAGCCAGGCTTCCGGGCCACCGGTCCC 259
QY 286 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGly 305
DB 260 CCTGGATTCCCAAGGCAAGTTGGATCACCTGGCCCACTGGCCCTCAAGCAGAGAAGGC 319
QY 306 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyPro 324
DB 320 AGCGAAGGATTCGAGGCCCCATCAGGCGCTGCCTGGCTCCCTGGGCCACCGGACCTCT 379
QY 325 -----GlnGlyProSer----- 328
DB 380 GGGATTGAGGCCCCCGCGTCTGGATGTTGGATGGGAGGATGGCAAGCTGGCTTG 439
QY 329 ---GlyAspProGlyProGlyProPro----- 337
DB 440 AGGGGGGACCTGGTCTCTGCTGGCCCCCTGGACTCATGGGACCACCGGGCTTAAGGG 499
QY 338 -----GlyLysGluGlyLeuProGln-----GlyProPro 349
DB 500 AAAACAGGACATCTCTGGCCCTCCAGGACCTAAGGGTGACTGGGCAAAACAGGCTCTCT 559
QY 350 GlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeu 369
DB 560 GGCAGCACTGGCGCGCTGGCCAGAGGCTGAACCTGGTGGCCATGGGACCCAGGGAAGA 619

Alignment Scores:
Pred. No.: 3,99e-17 Length: 1133
Score: 455.00 Matches: 98
Percent Similarity: 47.46% Conservative: 14
Best Local Similarity: 41.53% Mismatches: 50
Query Match: 15.33% Indels: 74
DB: 9 Gaps: 7

US-09-763-712A-2 (1-547) x US-09-992-600A-57 (1-1133)
QY 246 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 265
DB 140 CTCCAGGCCCCCAGGCCCCAGGCGTCCAGAGGAGCGAGGACCCCAAGGTAACCTCC 199
QY 266 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 285
DB 200 GGTGAGAAGGGCCACCGAGGATTCAGGCGCCAGCCAGGCTTCCGGGCCACCGGTCCC 259
QY 286 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGly 305
DB 260 CCTGGATTCCCAAGGCAAGTTGGATCACCTGGCCCACTGGCCCTCAAGCAGAGAAGGC 319
QY 306 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyPro 324
DB 320 AGCGAAGGATTCGAGGCCCCATCAGGCGCTGCCTGGCTCCCTGGGCCACCGGACCTCT 379
QY 325 -----GlnGlyProSer----- 328
DB 380 GGGATTGAGGCCCCCGCGTCTGGATGTTGGATGGGAGGATGGCAAGCTGGCTTG 439
QY 329 ---GlyAspProGlyProGlyProPro----- 337
DB 440 AGGGGGGACCTGGTCTCTGCTGGCCCCCTGGACTCATGGGACCACCGGGCTTAAGGG 499
QY 338 -----GlyLysGluGlyLeuProGln-----GlyProPro 349
DB 500 AAAACAGGACATCTCTGGCCCTCCAGGACCTAAGGGTGACTGGGCAAAACAGGCTCTCT 559
QY 350 GlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeu 369
DB 560 GGCAGCACTGGCGCGCTGGCCAGAGGCTGAACCTGGTGGCCATGGGACCCAGGGAAGA 619
QY 370 GlyLeuPro----- 373
DB 620 CCGGGTCCCCGGGACACGTGGGCCACCAGGCGCTCCAGGCGCCAGCCAGGACGCTGG 679
QY 374 -----Gly 374
DB 680 ATCTCTGCACTGGGTCTGAAGGAGACCGAGGAGCCACCGGAGAAAGGGCCCTTGCAGGC 739
QY 375 ValProGlyMetProGlyProLysGlyProGlyProGlyProPro-----GlyProSerGly 392
DB 740 CTCCAGGCGACCGCGGCCCGCCAGGCTCTCAAGGTCTCTCAGGCTATGGCAAGATGGGT 799
QY 393 AlaValValProLeuAlaLeuGlnAsnGluPro---ThrProAlaPro 407
DB 800 GCACAGGACCATGGGCGCAGCAAGCATCTCCCTGGCATCCCTGGGCC 847

RESULT 4
US-09-992-600A-57
; Sequence 57, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US4.DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
```


; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 782
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-782

Alignment Scores:
Pred. No.: 2.77e-16 Length: 6728
Score: 452.50 Matches: 99
Percent Similarity: 52.11% Conservative: 12
Best Local Similarity: 46.48% Mismatches: 53
Query Match: 15.24% Indels: 49
DB: 10 Gaps: 8

US-09-763-712A-2 (1-547) x US-09-954-456-782 (1-6728)

Qy 246 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 265
:::|||||
Db 1485 GTTCAAGGACCCCTGGCCCTCTGGAGAGGAAGAAAGCGAGGAGCTCGAGGTGAACCC 1544
Qy 266 GlyProThrGlyAsnLysGlnGlyGlyGluLysGlyGluPro----- 280
|||||
Db 1545 GGACCCACTGCGCTCCCGGACCCCTGGCGAGGCTGGTGACCTGGTAGCGGTGTTTC 1604
Qy 281 -----GlyProGlyProAlaGlyGluArgGlyProIleGly 293
|||||
Db 1605 CCTGCGCAGATGGTGTGGTCCCAAGGTCCTGGTGAAGCTGGTGAAGCTGGTTCCTCGC 1664
Qy 294 ProAlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGlySerGln 310
|||||
Db 1665 CCCGCTGGCCCAAGGATCTCTGGTGAAGCTGGTCCCGGTAAGCTGGTTCGCT 1724
Qy 311 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyThrValGlyGluPro 330
|||||
Db 1665 CCCGCTGGCCCAAGGATCTCTGGTGAAGCTGGTCCCGGTAAGCTGGTTCGCT 1724
Qy 350 -----GlyPheGlnGlyLeuGlnGlyThrValGlyGluPro 361
|||||
Db 1836 GCCCGTGGTCAAGGCTGGTGTGATGGGATTCCTGGACCTAAAGCTGGTGGAGAGGCC 1895
Qy 362 GlyValProGlyProGlyLeuProGlyLeuProGlyVal----- 375
|||||
Db 1895 GCGAAGCTGGAGAGGAGGTGTCCCGGACCCCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1955
Qy 376 ProGlyMetProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 392
|||||
Db 1956 GATGAGAGGCTGGAGCTCAGGAGCCCTGGCCCTGTGCTGGCTGGCTGGCTGGCTGG 2015
Qy 393 -----AlaValValProLeuAlaLeuGlnAsnGluProThrProAla----- 406

Db 2016 GAACAAGGCCCTGGTGGCTCCGCC---GGATTCCAGGGTCTCCCTGGTCTGCTGCTCT 2072
Qy 407 -----ProGluAspAsnGlyCysPro 413
|||||
Db 2073 CCAGGTGAAGCAGGCAACCTGGTGAACACAGGGGTGTTCT 2111
RESULT 8
US-09-880-107-3946
; Sequence 3946, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3946
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274615
US-09-880-107-3946

Alignment Scores:

Pred. No.: 2.77e-16 Length: 6728
Score: 452.50 Matches: 99
Percent Similarity: 52.11% Conservative: 12
Best Local Similarity: 46.48% Mismatches: 53
Query Match: 15.24% Indels: 49
DB: 10 Gaps: 8

US-09-763-712A-2 (1-547) x US-09-880-107-3946 (1-6728)

Qy 246 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 265
:::|||||
Db 1485 GTTCAAGGACCCCTGGCCCTCTGGAGAGGAAGAAAGCGAGGAGCTCGAGGTGAACCC 1544
Qy 266 GlyProThrGlyAsnLysGlnGlyGlyGluLysGlyGluPro----- 280
|||||
Db 1545 GGACCCACTGCGCTCCCGGACCCCTGGCGAGGCTGGTGACCTGGTAGCGGTGTTTC 1604
Qy 281 -----GlyProGlyProAlaGlyGluArgGlyProIleGly 293
|||||
Db 1605 CCTGCGCAGATGGTGTGGTCCCAAGGTCCTGGTGAAGCTGGTGAAGCTGGTTCCTCGC 1664
Qy 294 ProAlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGlySerGln 310
|||||
Db 1665 CCCGCTGGCCCAAGGATCTCTGGTGAAGCTGGTCCCGGTAAGCTGGTTCGCT 1724
Qy 311 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 330
|||||
Db 1725 GGTGCCAAGGCTGCTGACTGGAAGCCCTGGCAGCCCTGGTCTGATGGCAAACT----- 1778
Qy 331 ProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProProPro 349
|||||
Db 1779 ---GGCCCCCTGGTCCCGCGGTCAAGATGGTCCCGGACCCCGGAGGCCACCTGGT 1835
Qy 350 -----GlyPheGlnGlyLeuGlnGlyThrValGlyGluPro 361
|||||
Db 1836 GCCCGTGGTCAAGGCTGGTGTGATGGGATTCCTGGACCTAAAGCTGGTGGAGAGGCC 1895
Qy 362 GlyValProGlyProArgGlyLeuProGlyLeuProGlyVal----- 375


```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2094
; LENGTH: 5416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 J03464
US-09-880-107-2094

Alignment Scores:
Pred. No.: 6,53e-16 Length: 5416
Score: 444.00 Matches: 115
Percent Similarity: 43.40% Conservative: 23
Best Local Similarity: 36.16% Mismatches: 79
Query Match: 14.95% Indels: 102
DB: 10 Gaps: 10

US-09-763-712A-2 (1-547) x US-09-880-107-2094 (1-5416)

QY 197 IleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAsp 216
Db 2280 GTCCATTGGAAGCGAGGTCTCTTGGACCCCGCCAGGCGCTGATGAAACAAGGTGAAC 2339
QY 217 ThrGluValAlaAsnLeuSerValIleMetGluGluMetLysLeuValAspSer----- 234
Db 2340 CTGGTGTGGTGTGGTGGCAGCTGTGGTCCATCTGTCTAGTGACTCCAGGAG 2399
QY 235 -----LysHisGlyGlnLeuIleLysAsn 242
Db 2400 AGAGGGGTGCTCTGGCATACCTGTGAGGCAAGGAGAAAGGTGAACCTGTCTCAGAG 2459
QY 243 -----PheThrIleLeuGlnGlyPro 249
Db 2460 GTGAATTTGGTAACCTGSCAGAGATGTGTGCTGTGCTCATGTGCTGTA-GGTGCC 2518
QY 250 ProGlyProArgGlyProArgGlyAspArg-----GlySerGlnGlyProGly 266
Db 2519 CTTGTCTCTGTCGAGCCACAGAGTACCGGGCGGCAAGCTGGGGCTGCTGTCTGTGT 2578
QY 267 ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAla 286
Db 2579 CTGTGTGTCTCGGGGAAGCCCTGTGNACTGGCGAGGTGGTCTCTGCTGCCCCAC 2638
QY 287 GlyGluArgGlyProIleGlyProAlaGlyProPro-----GlyGluArgGlyGly 303
Db 2639 GGATTTGCTGTCGCGTGTCTGTCTGCTCAACCGGGTGTAAAGGAGAAAGAGAGGC 2698
QY 304 LysGlySerLysGlySerGln----- 310
Db 2699 AAAGGGCTTAAGGGTGAACACGGTGTGTGTGTCACAGAGCCCGCTTGGAGCTGTGTC 2758
QY 311 -----GlyProLysGlySerArgGlySer----- 318
Db 2759 CCAGCTGTGTCCAAATGGTCCCGCGTCTGTGTGAAGTGGTGTGATGGAGGCCCGCT 2818
QY 319 -----ProGlyLysProGlyProGlnGlyPro 327
Db 2819 GGTATGACTGGTTTCCCTGGTCTGCTGGAGCGACTGGTCCCGCCAGCCCTCTGGTAT 2878
QY 328 SerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGln--- 346
Db 2879 TCTGGCCCTCTGTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2938
QY 347 -----GlyProProGlyPheGln 352
Db 2939 GACCAAGTCTCAGTTGGCGCAACTGGAGAGTAGTGCATGGTGGTCCCGCTGCTGCT 2998
QY 353 GlyLeuGln-----GlyThrValGlyGluProGlyValProGlyPro 366
Db 2999 GTGTAGAAGGGTCCCTCTGAGAGGCTGTGTACTGTGGACCTCTCTGGCACTCCAGTCT 3058
QY 367 ArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGly 386
Db 3679 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3938
```

```
Db 3059 CAGGGTCTTTTGGTGTCTCTCTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3118
QY 387 ProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGluPro 403
Db 3119 CTACCTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3178
QY 404 -----ThrProAlaProGluAspAsnGlyCysPro 413
Db 3179 GGGGCGCGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3232

RESULT 11
US-10-044-090-104
; Sequence 104, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 104
; LENGTH: 4149
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1393093.19
; NAME/KEY: unsure
; LOCATION: 1492-1495, 1663, 1669, 4012, 4055
; OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-104

Alignment Scores:
Pred. No.: 7,92e-16 Length: 4149
Score: 440.50 Matches: 95
Percent Similarity: 49.32% Conservative: 13
Best Local Similarity: 43.38% Mismatches: 58
Query Match: 14.84% Indels: 53
DB: 12 Gaps: 6

US-09-763-712A-2 (1-547) x US-10-044-090-104 (1-4149)

QY 248 GlyProProArgGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 267
Db 1307 GTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1366
QY 268 Thr-----Gly 269
Db 1367 AATGGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1426
QY 270 AsnLysGlyGlnLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 289
Db 1427 GCCAAAGGCCCTTAAGGGTGAACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1486
QY 290 GlyProIleGlyProAlaGlyProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 309
Db 1487 GGCNNNNNGTCCAAATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1546
QY 310 GlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyPro----- 327
Db 1547 CCTGTATGACTGGTTTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1606
QY 328 ---SerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGln 346
Db 1607 ATTTCTGCCCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1666
QY 347 -----GlyProProGlyPhe 351
Db 1667 GNGACCAAGGACCAAGGACCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1726
QY 352 GlnGlyLeuGln-----GlyThrValGlyGluProGlyValProGly 365
```

```
Db 1727 GCTGGTGAAGGGTCCCTCTCTGGAGAGGCTGGTACTGCTGGACCTCTGGCACTCCAGGT 1786
QY 366 ProArgLeuProGlyLeuProGlyValProGlyMetProGlyProGlyProGlyPro 385
Db 1787 CTCAGGGTCTTCTGGTCTCTGGTATCTCTGGGCTCTCTGGGCTCTGGGCTCTGGGCTCT 1846
QY 386 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 402
Db 1847 GGTCTACCTGGTGTCTGGTCTGGTGTGAACCTGGTCTCTGGCAATTCGGGCCCT 1906
QY 403 Pro-----ThrProAlaProGluAspAsnGlyCysPro 413
Db 1907 CCTGGGCCCGTGGTCTCTCTGGTGTCTGGTGTAGTCTCTGGAGTCAACGGTCTCTCT 1963

RESULT 12
US-09-880-107-3947
; Sequence 3947, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3947
; LENGTH: 5086
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274616
; NAME/KEY: unsure
; LOCATION: (1)..(5086)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3947

Alignment Scores:
Pred. No.: 9,53e-16 Length: 5086
Score: 440.50 Matches: 114
Percent Similarity: 43.08% Conservatives: 23
Best Local Similarity: 35.85% Mismatches: 80
Query Match: 14.84% Indels: 102
DB: 10 Gaps: 10

US-09-763-712A-2 (1-547) x US-09-880-107-3947 (1-5086)
QY 197 IleArgLeuAspSerValSerLeuArgMetGlnAspLeuMetArgSerArgLeuAsp 216
Db 1950 GTCCATTGTGAAGCGGAGGTCTCTTGGACCCAGGCGCTCATGAAACAAGGGTGAAC 2009
QY 217 ThrGluValAlaAsnLeuSerValIleMetGluMetLysLeuValAspSer----- 234
Db 2010 CTGGTGTGGTGTGGTGGGCACTGCTGTCCATCTGTGGTCTCTAGTGGACTCCCGAGG 2069
QY 235 -----LysHisGlyGlnLeuIleLysAsn 242
Db 2070 AGAGGGGTCTCTGGCATACCTGGAGCAAGGAGAAAGGGTGAACCTGGTCTCAGAG 2129
QY 243 -----pheThrIleLeuGlnGlyPro 249
Db 2130 GTCAAAATTGGTAACCCCTGGCAGAGATGFGCTCGTGGTCTCATGCTGTGTA-GGTGCC 2188
QY 250 ProGlyProArgGlyProArgGlyAspArg-----GlySerGlnGlyProGlyPro 266
```

```
Db 2189 CCTGTCTCTCTCTGGAGCCACAGGTGACCGGGCGGAAGCTGGGGCTGCTGTCTCTGTGTT 2248
QY 267 ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAla 286
Db 2249 CCTGTGTGGTCTCTGGGGGAAGCCCTGGTGAACGTGGCGAGGTGGTCTCTGTGCTGCCCAAC 2308
QY 287 GlyGluArgGlyProIleGlyProAlaGlyProPro-----GlyGluArgGlyGly 303
Db 2309 GGATTGTGTGGTCCGGCTGCTGCTGCTGCTCAACCGGGTGTAAAGAGAGAAAGAGAGGCC 2368
QY 304 LysGlySerLysGlySerGln----- 310
Db 2369 AAAGGGCTTAAGGGTCAAAACGGTGTGTGTGTGTCCACAGGCCCTGTGGAGCTGTGGC 2428
QY 311 -----GlyProLysGlySerArgGlySer----- 318
Db 2429 CCAGCTGGTCCAATAGTCCCCCGTCTCTGTGAAGTCGTGGTATGGAGGCCCTCT 2488
QY 319 -----ProGlyLysProGlyProGlnGlyPro 327
Db 2489 GGTATGACTGTTTCCCTGCTGCTGCTGGACGAGTGTCTCCCGAGACCTCTGTGATT 2548
QY 328 SerGlyAspProGlyProProGlyProGlyLysGluGlyLeuProGlyProGln---- 346
Db 2549 TCTGGCCCTCTCTGGTCTCTGGTCTCTGGTCTCTGGTCTCTGGTCTCTGGTCTCTGGT 2608
QY 347 -----GlyProProGlyPheGln 352
Db 2609 GACCAAGTCCAGTGGCGGAACGTGGAAGTAGTGTGACGTGGTCTCTCTCTCTCTCTCT 2668
QY 353 GlyLeuGln-----GlyThrValGlyGluProGlyValProGlyPro 366
Db 2669 GGTGAAGGGTCCCTCTCTGGAGAGGCTGTACTGTGGACCTCTCTGGCACTCCAGTCT 2728
QY 367 ArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGly 386
Db 2729 CAGGCTCTCTGTGGTCTCTCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2788
QY 387 ProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGluPro 403
Db 2789 CTACCTGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2848
QY 404 -----ThrProAlaProGluAspAsnGlyCysPro 413
Db 2849 GGGGCCCGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2902

RESULT 13
US-09-925-299-206
; Sequence 206, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
```



```
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Alignment Scores:
Pred. No.: 9.63e-16 Length: 5145
Score: 440.50 Matches: 95
Percent Similarity: 49.32% Conservative: 13
Best Local Similarity: 43.38% Mismatches: 58
Query Match: 14.84% Indels: 53
DB: 9 Gaps: 6

US-09-763-712A-2 (1-547) x US-09-925-299-206 (1-5145)
Qy 248 GlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 267
||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 2248 GGTCTGTGTCCTCGGGGAAAGCCCTGGTGAACGTGGTGGTGGTGGTGGTGG 2307
Qy 268 Thr-----Gly 269
Db 2308 AATGATTTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTC 2367
Qy 270 AsnLysGlyGlnLysGlyGluLysGlyProGlyProGlyProGlyProGly 289
||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 2368 GCCAAGGCGCTAAGGTGAAACGGTGTGTTGTCCTCCAGGCGCCCTGGAGT 2427
Qy 290 GlyProLleGlyProAlaGlyProProGlyGluArgGlyGlySerLysGly 309
||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 2428 GGCCAGCTGTCTCAATAGTGTCTGCTGGAAGTGGTGGTGGTGGTGGTGG 2487
Qy 310 GlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGly 327
||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 2488 CTGATGATGACTGTTTCCCTGGTCTGCTGGAGAGGCTGGTGGTGGTGGT 2547
Qy 328 ---SerGlyAspProGlyProGlyProGlyProGlyProGlyProGlyPro 346
||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 2548 ATTTCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2607
Qy 347 -----GlyProGlyPhe 351
Db 2608 GGTGACCAAGTCCAGTTGGCCGAAGTAGTGGTGGTGGTGGTGGTGGTGG 2667
Qy 352 GlnGlyLeuGln-----GlyThrValGlyGluProGlyValProGly 365
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2668 GCTGTGAGAGGGTCCCTCTGGAGAGGCTGGTACTGTGGACCTCTCCGAC 2727
Qy 366 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProGly 385
||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 2728 CCTCAGGCTCTTCTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2787
Qy 386 GlyProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGln 402
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2788 GGTCTACCAAGTGTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2847

RESULT 14
US-09-925-299-206
; Sequence 206, Application US/09925299
; Patent No. US2002005562A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P4102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:08:41 ; Search time 1808.03 Seconds
(without alignments)
4899.781 Million cell updates/sec

Title: US-09-763-712A-2
Perfect score: 2969
Sequence: 1 MYSHNVIMNLNLTLQVQ.....EDVNNFICKDRETVLSSAL 547

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n_model -DEV=xlp
-Q=/cgn2_1/USPRO_pool/US09763712/runat_14032003_100950_18109/app_query.fasta_1.1877
-DB=EST -QFMT=fastap -SUFMT=first -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712 -CGN_1_1_4575 -runat_14032003_100950_18109 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_estl.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|----------|--------------------|
| 1 | 2765 | 93.1 | 3305 | 11 | BC009162 | BC009162 Mus muscu |
| 2 | 1460.5 | 49.2 | 957 | 14 | BQ955927 | BQ955927 AGENCOURT |
| 3 | 1394.5 | 47.0 | 936 | 14 | BQ891432 | BQ891432 AGENCOURT |
| 4 | 1256 | 42.3 | 861 | 14 | BQ713873 | BQ713873 AGENCOURT |
| 5 | 1251 | 42.1 | 906 | 13 | BI456109 | BI456109 603172765 |
| 6 | 1219.5 | 41.1 | 884 | 14 | BQ934501 | BQ934501 AGENCOURT |
| 7 | 1022 | 34.4 | 580 | 12 | BE910803 | BE910803 601661855 |
| 8 | 1007 | 33.9 | 668 | 14 | BQ674807 | BQ674807 AGENCOURT |
| 9 | 957 | 32.2 | 601 | 10 | BE290299 | BE290299 601089246 |
| c 10 | 916.5 | 30.9 | 967 | 9 | AL568743 | AL568743 AL568743 |
| 11 | 901 | 30.3 | 581 | 14 | BQ127513 | BQ127513 1160h06.y |
| 12 | 842 | 28.4 | 552 | 10 | AW958053 | AW958053 EST370123 |
| 13 | 831 | 28.0 | 638 | 10 | BB248064 | BB248064 BB248064 |
| c 14 | 777.5 | 26.2 | 511 | 12 | BF805750 | BF805750 QV1-C1017 |
| 15 | 767 | 25.8 | 808 | 14 | BQ771366 | BQ771366 UI-M-F10- |
| c 16 | 669 | 22.5 | 500 | 14 | BM676508 | BM676508 UI-E-EJ0- |
| 17 | 663 | 22.3 | 507 | 14 | BM713891 | BM713891 UI-E-EJ0- |
| c 18 | 650 | 21.9 | 410 | 12 | BF856923 | BF856923 QV1-FT020 |
| 19 | 627.5 | 21.1 | 591 | 13 | BI445884 | BI445884 dal33d10. |
| 20 | 602 | 20.3 | 1082 | 14 | BM907108 | BM907108 AGENCOURT |
| 21 | 600 | 20.2 | 357 | 10 | AW379436 | AW379436 CM4-HT024 |
| 22 | 594 | 20.0 | 350 | 9 | AA012704 | AA012704 RPUL402CG |
| 23 | 588 | 19.8 | 339 | 9 | AA304251 | AA304251 EST17006 |
| c 24 | 567 | 18.9 | 517 | 9 | AL543000 | AL543000 AL543000 |
| 25 | 530 | 18.1 | 310 | 9 | AA361740 | AA361740 EST71069 |
| 26 | 516 | 17.4 | 465 | 10 | AW240221 | AW240221 up30b07.y |
| 27 | 503.5 | 17.0 | 839 | 13 | BI697412 | BI697412 603348154 |
| 28 | 478 | 16.1 | 943 | 14 | BQ887163 | BQ887163 AGENCOURT |
| c 29 | 477 | 16.1 | 679 | 14 | W27610 | W27610 35b6 Human |
| 30 | 460 | 15.5 | 441 | 10 | BB749743 | BB749743 BB749743 |
| 31 | 458.5 | 15.4 | 569 | 10 | AV616076 | AV616076 AV616076 |
| 32 | 457.5 | 15.4 | 1029 | 17 | CNS04VYG | AL309897 Tetraodon |
| 33 | 456.5 | 15.4 | 792 | 13 | BI456015 | BI456015 603170460 |
| 34 | 456 | 15.4 | 420 | 12 | BE929813 | BE929813 RC3-GN004 |
| 35 | 455.5 | 15.3 | 896 | 14 | BQ955616 | BQ955616 AGENCOURT |
| 36 | 453.5 | 15.3 | 877 | 12 | BG681943 | BG681943 602629884 |
| 37 | 452 | 15.2 | 688 | 13 | BI249235 | BI249235 602995511 |
| 38 | 452 | 15.2 | 3951 | 11 | BC013626 | BC013626 Mus muscu |
| 39 | 451.5 | 15.2 | 921 | 14 | BQ921102 | BQ921102 AGENCOURT |
| 40 | 451.5 | 15.2 | 1192 | 14 | BM905999 | BM905999 AGENCOURT |
| 41 | 449.5 | 15.1 | 907 | 14 | BQ222937 | BQ222937 AGENCOURT |
| 42 | 448 | 15.1 | 946 | 14 | BQ942234 | BQ942234 AGENCOURT |
| 43 | 447 | 15.1 | 862 | 12 | BG171919 | BG171919 602333020 |
| 44 | 446.5 | 15.0 | 608 | 13 | BM425818 | BM425818 pgf2c.Bk0 |
| 45 | 446 | 15.0 | 651 | 10 | BB613976 | BB613976 BB613976 |

ALIGNMENTS

RESULT 1
BC009162
LOCUS BC009162 3305 bp mRNA linear HTC 07-AUG-2002
DEFINITION Mus musculus, Similar to scavenger receptor with C-type lectin,
clone IMAGE:2811487, mRNA.
ACCESSION BC009162
VERSION BC009162.1 GI:14714370
KEYWORDS HTC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3305)
AUTHORS Strausberg, R.
TITLE Direct Submission

JOURNAL

Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs@mail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 5 Row: h Column: 10
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis

This clone has the following problem: incomplete processing.

FEATURES
Source

Location/Qualifiers
 1..3305
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="C57BL/6J"
 /clone="IMAGE:2811487"
 /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."
 /clone_lib="NCI CGAP_Mam5"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 BASE COUNT 1031 a 757 c 760 g 757 t

Alignment Scores:

Pred. No.: 4.24e-181 Length: 3305
 Score: 2765.00 Matches: 502
 Percent Similarity: 96.71% Conservative: 27
 Best Local Similarity: 91.77% Mismatches: 18
 Query Match: 93.13% Indels: 0
 DB: 11 Gaps: 0

US-09-763-712A-2 (1-547) x BC009162 (1-3305)

QY 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnAsnLeuThrGlnValGln 20
 ||||| :|||
 DB 681 ATGATTCAGAGTGTGGTTATCATGAACCTCAACAACTGAACCTGACCCAGGTTTCAG 740
 QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
 ||||| :|||
 DB 741 CAGAGGAACCTTATCTCAAAATTTGCAGCAGTCTGTGGATGACACAAGCCCTGGCCATCCAG 800
 QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnGlnValPheLeuGlnAlaLysLysAspThr 60
 ||||| :|||
 DB 801 CGAATTAAGAAGATTTCAAAATCTGCAGCAGGTTTTCTTCAAGCCCAAGAGGACACC 860
 QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
 ||||| :|||
 DB 861 GATTGGCTAAAGGAAAAGTACAAAGCTTGCAGACATTGGCTGCCAACAACTCTGCCCTG 920
 QY 81 AlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
 ||||| :|||
 DB 921 GCCAAAGCCCAACAACTGACACCTTAGAGGATATGAATAGCCAGCTCATTCACAGT 980
 QY 101 GlnMetGluAsnIleThrIleSerGlnAlaAsnGlnAsnLeuLysAspLeuGln 120
 ||||| :|||
 DB 981 CAGATGGACAACATTAACCACTATCTCACAGGGCCACGACGAGCCTGAAGACCTTCAG 1040

QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluGluArg 140
 ||||| :|||
 DB 1041 GACTTACACAAGGATACAGAAATACAGAGCTGTCAAGTTCAAGCCAACTTGAGGAACGC 1100
 QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerThrAlaHis 160
 ||||| :|||
 DB 1101 TTTCAGGTCTTTGAGACAGATATTGTGAACATCATTTAGCAACATCAGCTACACAGCCCAT 1160
 QY 161 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 180
 ||||| :|||
 DB 1161 CATCTCGGACACTGACACGAACTCTGAATGATGTTAGGACACATGACAGACACTTG 1220
 QY 181 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 200
 ||||| :|||
 DB 1221 ACCAGACACAGGATGACCTGACCTCTCTTGAATAACACACTAGTCAACATCCGCTGGAT 1280
 QY 201 SerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAla 220
 ||||| :|||
 DB 1281 TCTATTCTCTCAGGATGCGCAAGACATGATGAGTCAAAAGTTAGACACTGAAGTGGCC 1340
 QY 221 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIle 240
 ||||| :|||
 DB 1341 AACTTATCAGTGGTTATGGAAGAGATGAACCTGTTGACTTCCAAGCAGGTCACTCATC 1400
 QY 241 LysAsnPheThrIleLeuGlnGlyProGlyProGlyProArgGlyProArgGlyAspArgGly 260
 ||||| :|||
 DB 1401 AAGAACCTTTACCATTCTACAAAGGTCTCTCTGGCCCGCAGAGGTCCAAAGGTGACAGGA 1460
 QY 261 SerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro 280
 ||||| :|||
 DB 1461 TCTCAGGGACCACTCGTCCAACTGGCAACAAGGGGACAGAAAGGAGAGGAGAGCCCT 1520
 QY 281 GlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlu 300
 ||||| :|||
 DB 1521 GGTCCACCTGGCCCTCGGGTGAGAGGGGCACAAATTTGGACCAGTCCGCCCTCTCTGGAGAG 1580
 QY 301 ArgGlyGlyLysGlySerLysGlySerGlnGlyProGlyProLysGlySerArgGlySerProGly 320
 ||||| :|||
 DB 1581 CGTGGCAGCAAAAGGATCCAAAGGCTCACAGGGTCCCAAGGATCTCGTGGTCCCGAGG 1640
 QY 321 LysProGlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGlu 340
 ||||| :|||
 DB 1641 AAGCCTGGCCCTCAAGGACCTAGTGGGGACCCAGGACCCAGGTCACACAGCAAGGAT 1700
 QY 341 GlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 360
 ||||| :|||
 DB 1701 GGACTCCCTGGCCCTCAGGGCCCTCTCTGGCTTCCAGGGACTACAGGGCAGTGTGGGTGAG 1760
 QY 361 ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGly 380
 ||||| :|||
 DB 1761 CCTGGAGTACCTGGACCTCGGGGTTGCCAGGCTTCCAGGGTGCAGGCATGCCCTGG 1820
 QY 381 ProLysGlyProProGlyProProGlyProSerGlyValAlaValProLeuAlaLeuGln 400
 ||||| :|||
 DB 1821 CTTAAGGGACCACTGGCCCTCCAGGCCCTCAGGAGCAATGGAGCATTTGGCTCTGCAG 1880
 QY 401 AsnGluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThr 420
 ||||| :|||
 DB 1881 AATGAACCAACCCAGCATCAGAGGTCAACGGATGTCCACCTCAGTGGAGAACTTCACA 1940
 QY 421 AspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 440
 ||||| :|||
 DB 1941 GATAAATGCTACTATTTTTCATTGGAAAAAGAAATTTTGAAGATGCTAAGCTTTTCTGT 2000
 QY 441 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 460
 ||||| :|||
 DB 2001 GAAGCAAAATCTTCCCATCTCGTTTTCATAAACTCAAGAGAGAACAGCAATGGATATAA 2060
 QY 461 LysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 480
 ||||| :|||
 DB 2061 AAGCATACCGTGGGAGAGAAAGCCATTGGATCGGCCCTCACAGACTCACAGCAGGAAGC 2120

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman, Ph.D.
CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13790 row: n column: 19
High quality sequence start: 25
High quality sequence stop: 632.
Location/Qualifiers
1..936
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6332754"
/clone_lib="NIH_MGC_130"
/lab_host="DH10B (phage-resistant)"
/note="Organ: oocytes; Vector: pCMV-SPORT6.1.ccd; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.95 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
BASE COUNT 273 a 250 c 244 g 163 t 6 others
ORIGIN
Alignment Scores:
Pred. No.: 5,71e-87 Length: 936
Score: 1394.50 Matches: 272
Percent Similarity: 95.05% Conservative: 16
Best Local Similarity: 89.77% Mismatches: 9
Query Match: 46.97% Indels: 6
DB: 14 Gaps: 1
US-09-763-712a-2 (1-547) x B0891432 (1-936)
QY 60 ThrAspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnSerAla 79
|||||
Db 38 ACCGATTGGCTAAGGAAGAAGTACAGAGCTTCGACACATTTGGCTGCCAACAACTCTGCC 97
QY 80 LeuAlaLysAlaAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThr 99
|||||
Db 98 CTGGCCAAAGCCCAATGACACCCCTAGAGGATATGAATGACCGAGCTCAGCTCATTCACA 157
QY 100 GlyGlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeu 119
|||||
Db 158 GGTACATGGACACAACTATACCACTATCTCACAGGCCCAAGCAGCAGACCTGAAAGACCTT 217
QY 120 GlnAspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluGlu 139
|||||
Db 218 CAGACTTACCAAGGATACAGAAATAGAACAGCTGTCAAGTTCAGCCAACTTGAGGAA 277
QY 140 ArgPheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAla 159
|||||
Db 278 CGCTTCAGGCTCTTCAGACAGATATTGTGAACATCATTTAGCAACATCAGCTACACAGCC 337
QY 160 HisHisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThr 179
|||||
Db 338 CATCACCTGAGGACACTGACCAACATCTGAATGTATGTAGGACCATGACAGACACACC 397
QY 180 LeuThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAlaAsnIleArgLeu 199
|||||
Db 398 TTGACCAAGACACAGGATGACCTGACCTCTTGAATTAACACACTAGTCAACATCCGCTTG 457
QY 200 AspSerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAspThrGluVal 219
|||||
Db 458 GATTCTATTCTCTCAGGATGAGCAAGACATCATGAGCTCAAGTTCAGACACTGAAGTG 517
QY 220 AlaAsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeu 239
|||||

Db 518 GCCAACTTATCAGTGGTATTGGAAGAGATGAAACTGGTTGACTCCAGCAGCGTCAGCTC 577
QY 240 IleLysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyProArgGlyAspArg 259
Db 578 ATCAAGAACTTTTACCATTCTACAAAGGTCTCTGGCCCCCAGAGGTCCAAAGGTGACAGA 637
QY 260 GlySerGlnGlyProGlyProArgGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGlu 279
Db 638 GGATCTCAGGACCACTGGTCCAACTGGCAACAAGGACAGAAAGAGAGAGAGAGAG 697
QY 280 ProGlyProGlyProAlaGlyGluArgGlyProIleGlyProIleGlyProAlaGlyProGly 299
Db 698 CTTGGTCCACTGGCCCTGGGGTGAGAGGGGCACAAATTGGACCACTCGGCCCTCCTGGA 757
QY 300 GluArgGlyGlyLysGlySerLysGlySerGlnGly-ProLysGlySerArgGlySer-P 319
Db 758 GAGCGTGGCAGCAAGGATCCAAAGGCTCACAGGGNTCCAAAGGGATCTCGTGGGTCCCC 817
QY 319 roGlyLysProGlyProGlnGlyProSerGlyAspProGly-ProProGlyProProGly 338
Db 818 CAGGGAAGCTGGCCCTCAAGGAGCTAGTGGGACCCAGGACCCAGGACCCAGGACCCAGG 877
QY 339 -LysGluGlyLeuPro-GlyProGlnGlyProPro---GlyPheGlnGlyLeuGlnGly 356
Db 878 CAGGGATGAGTCCCTGGGCCCTCAGGGNCNCNTCTCTGGGCTCNCAGGGGACTACAAGG 936
RESULT 4
BQ713873 861 bp mRNA linear EST 16-JUL-2002
LOCUS AGENCOURT_8482828 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6306240
DEFINITION 5', mRNA sequence.
ACCESSION BQ713873
VERSION BQ713873.1 GI:21852772
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 861)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13721 row: n column: 01
High quality sequence stop: 598.
Location/Qualifiers
1..861
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6306240"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1.ccd; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
BASE COUNT 221 a 253 c 239 g 147 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 1.94e-77 Length: 861
Score: 1256.00 Matches: 229
Percent Similarity: 91.01% Conservative: 14

| | | | | | | |
|------------|---|-------------|---------------|--------|--------|-----------------|
| RESULT 5 | BI456109 | 603172765F1 | NCI_CGAP_Mam5 | 906 bp | linear | EST 21-AUG-2001 |
| LOCUS | BI456109 | | | | | |
| DEFINITION | Mus musculus cdna clone IMAGE:5251888 5', mRNA sequence. | | | | | |
| ACCESSION | BI456109 | | | | | |
| VERSION | BI456109.1 | | | | | |
| KEYWORDS | GI:15246765 | | | | | |
| SOURCE | house mouse. | | | | | |
| ORGANISM | Mus musculus | | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | | |

```
Db 423 TCCAAAGGCTCAGAGGTCCTCCAAAGGATCTCGTGGTCCCGAGTGAAGCTGGCCCTCAA 482
QY 326 GlyProSerGlyAspProGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 345
Db 483 GGACCTAGTGGGACCCAGGACCCAGGTCACAGGTCACAGGTCACAGGTCACAGGTCAC 542
QY 346 GlnGlyProGlyPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 365
Db 543 CAGGGCCCTCCTGGCTTCCAGGACTACAGGACATGTGGGTGAGCTGGAGTACCTGGA 602
QY 366 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProGlyProGlyPro 385
Db 603 CCTCGGGGTTGCCAGGCTTCCAGGGTCCAGGATGCTGGGCTTAAGGACCACT - 661
QY 386 GlyProGlyProSerGlyAlaValValProLeuLeuAlaLeuGlnAsnGluProThrPro 405
Db 662 AGCCCTCCAGGCCCCCAGGAGATG - GAGCCATTCGTCGAGAATTAAGCAAGCCCA 720
QY 406 AlaProGlu - AspAsnGlyCysProProHisTrpLysAsnPheThrAsp - - - LysCysT 424
Db 721 GCATCAGAGGTTCAACGCGATGTCCTCTCACTGGAAGAACTTTTCACGGATTATGGCT 780
QY 424 yrTyPheSerVal - - - GluLysGluIlePheGluAspAlaLysLeuPheCysGluAspL 443
Db 781 ACTATTTTTCATTCACAAAAGAAATTTTGAAGATTGCTAAGCGTTCTGTGGAGAAA 840
QY 443 ysSerSerHis 446
Db 841 AATTTTCCAAT 851

RESULT 6
LOCUS BQ934501
DEFINITION AGENCOURT_8765685 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6330794
5' mRNA sequence.
ACCESSION BQ934501
VERSION BQ934501.1 GI:22349884
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 884)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13785 row: m column: 03
High quality sequence stop: 352.
FEATURES
Location/Qualifiers
1..884
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6330794"
/clone_lib="NIH_MGC_130"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: otocysts; Vector: pCMV-SPORT6.1.cdb;
Site:1: EcoRV; Site:2: NotI; Cloned unidirectionally.
Primer: Oligo dt. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a
NIH_MGC Library."
BASE COUNT 282 a 225 c 205 g 172 t
ORIGIN
..
```

```
Alignment Scores: 6,64e-75 Length: 884
Pred. No.: 1219,50 Matches: 255
Score: 92,54% Conservativeness: 18
Percent Similarity: 86,44% Mismatches: 14
Best Local Similarity: 41,07% Indels: 9
Query Match: 14 Gaps: 4
DB:
US-09-763-712a-2 (1-547) x BQ934501 (1-884)
QY 1 MetTySerHisAsnValValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
Db 7 ATGTATTCTCAGAGCGTGGTTATCATGAACCTCAACAACCTGAACCTCAACCCAGGTTCCAG 66
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
Db 67 CAGAGGAACCTTATCTCAAAATCTGCAGCAGTCTGTGGATGACACAGCCTGGCCATCCAG 126
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
Db 127 CGAATTAAGAATGATTCCAAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGGACACC 186
QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
Db 187 GATTGGCTAAAGGAAAAAGTACAGAGCTTGACAGACATTGGCTGCCAACAACCTGCCCCTG 246
QY 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThr-Gl 100
Db 247 GCCAAGGCCAACAATGACACCTAGAGGATATGAATAGCCAGCTCAGCTCATTTACAGGG 306
QY 100 yGlnMetGluAsnIleThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGl 120
Db 307 TCAGATGGACAACATTACCACTATCTCACAGGCCAAGCAGAGAGCCTGAAAGACCTTCA 366
QY 120 nAspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluAla 140
Db 367 GGACTTACACAAGGATACAGAAAATAGAACAGCTGTCTCAAGTTTCAGCCAACTTGAGGAG 426
QY 140 gPheGlnLeuPheGluThrAspIleValIleSerAsnIleSerAsnIleSerTyThrAlaH 160
Db 427 CTTCCAGGCTTTTGAGACAGATATTGTGAACATCATATTGACACATCAGCTACACAGCC 486
QY 160 shLysLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrCysThrAspThrLe 180
Db 487 TCACCTGAGGACACTGACCAAGCAATCTGAATGATGTAGGACCAACATGACACACACCTT 546
QY 180 uThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAs 200
Db 547 GACCAGACACAGGATGACCTGACCTCTTGAATAC-ACACATAGTCAACATCCGTTGGA 605
QY 200 pSerValSerLeuArgMetGlnAsnLeuMetArgSerArgLeuAspThrGluValAl 220
Db 606 TTTCTATTCTCTCAGAGTGACCAAGACATGATGAGGTCAAGTTAGACACTGAAGTGGC 665
QY 220 aAsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuI 240
Db 666 CAACCTTATCAGTGGTTATGGAAGAGATGAACTGGTTGACTCCAAGCAGCGTCACTCAT 725
QY 240 eLysAsnPheThrIleLeuGlnGlyProGlyProArgGly-ProArgGlyAspArgG 260
Db 726 CAAGAACCTTTACCTTCTACAAGGTCTCTCTGGCCCCAGAGGTCAAAAGGGGGACAGAG 785
QY 260 LysSerGlnGlyProPro - - - GlyProThr-GlyAsnLysGlyGlnLys - - - - - GlyGlu 276
Db 786 GGATCTCAGGGGACCACTGGGTCCCACTTGGCACACAGGGGGAACAAAAAGGGAGAGAA 845
QY 277 LysGlyGluPro - - - GlyProPro - - - GlyProAlaGly 287
Db 846 AGGGGAGAGAGCCTGTGGTCCACCTGGGGCCCTGGCGGC 884
RESULT 7
BE910803
```



```

LOCUS      BE910803      580 bp      mRNA      linear      EST 29-SEP-2000
DEFINITION 601661855F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3962292 5',
            mRNA sequence.
ACCESSION  BE910803
VERSION    BE910803.1  GI:10407765
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 580)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Inyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLM9128 row: m column: 13
            High quality sequence stop: 580.
            Location/Qualifiers
                Location: 1..580
                /organism="Mus musculus"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="IMAGE:3962292"
                /clone_lib="NCI_CGAP_Mam1"
                /tissue_type="tumor, biopsy sample"
                /dev_stage="10 months, virgin"
                /lab_host="DH10B"
            /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
            Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
            Library constructed by Life Technologies. Investigator
            providing samples: Gilbert Smith, NIH"
BASE COUNT 159 a 153 c 161 g 107 t
ORIGIN
Alignment Scores:
Pred. No.: 1.9e-61      Length: 580
Score: 1022.00      Matches: 177
Percent Similarity: 96.34%      Conservative: 7
Best Local Similarity: 92.67%      Mismatches: 0
Query Match: 34.42%      Indels: 0
DB: 12      Gaps: 0
US-09-763-712A-2 (1-547) x BE910803 (1-580)
QY 304 LysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGly 323
Db 6 AAAGATCCAAAGGTCACAGGGTCCCAAGGATCTCGTGGTCCCGAGGGAAGCTGGC 65
QY 324 ProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuPro 343
Db 66 CCTCAGGACCTAGTGGGGACCCAGACACAGGTCCACAGGCAAGATGGATCCCT 125
QY 344 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 363
Db 126 GGCCCTCAGGGGCCCTCTGGCTTCAGGGACTACAGGGCACGTGGGTGAGCCTGGAGTA 185
QY 364 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 383
Db 186 CTTGACCTCGGGGGTTCGACGGCTTGCAGGGGTGCCAGGGATGCCAGGATCGCTGGGCTAAGGGA 245
QY 384 ProProGlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 403
Db 246 CCACCTGGCCCTCCAGGCCCTCCAGAGCAATGGAGCACTGGCTCGCAGATGAACCA 305
QY 404 ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCys 423

```

```

Db 306 ACCCCAGCATCAGAGGTCAACGGATGTCGCTCACTGGAAGAACTTCACAGATAATGC 365
QY 424 TyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLys 443
Db 366 TACTATTTTTCATTGGAAAAAGAAATTTTGAAGATGCTGAAGCTTTCTGTGAAGACAA 425
QY 444 SerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMet 463
Db 426 TCTTCCCTCTCGTTTTCATAAATCAAGAGAAGAACAGCAATGGATAAAAAAGCATACC 485
QY 464 ValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLys 483
Db 486 GTGGGGAGAGAAGCAATGGATCGGCTCACAGACTCAAGAACAGAAAGCAATGAAG 545
QY 484 TrpLeuAspGlyThrSerProAspTyrLysAsn 494
Db 546 TGGCTAGACGGGTCACTGTTGATTACAAAAAC 578
RESULT 8
LOCUS      BQ674807      668 bp      mRNA      linear      EST 15-JUL-2002
DEFINITION AGENCOURT_8035172 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6212672
            5', mRNA sequence.
ACCESSION  BQ674807
VERSION    BQ674807.1  GI:21785641
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 668)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLM2377 row: c column: 09
            High quality sequence stop: 656.
            Location/Qualifiers
                Location: 1..668
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:6212672"
                /clone_lib="NIH_MGC_102"
                /tissue_type="epidermoid carcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                Site_2: EcoRI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Library constructed
                by Ling Hong in the laboratory of Gerald M. Rubin
                (University of California, Berkeley) using ZAP-CDNA
                synthesis kit (Stratagene) and Superscript II RT (Life
                Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 216 a 148 c 167 g 137 t
ORIGIN
Alignment Scores:
Pred. No.: 2.4e-60      Length: 668
Score: 1007.00      Matches: 174
Percent Similarity: 99.43%      Conservative: 0
Best Local Similarity: 99.43%      Mismatches: 1
Query Match: 33.92%      Indels: 0
DB: 14      Gaps: 0

```


ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 967)
AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
Location/Qualifiers
1..967
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DE005YH04"
/clone_lib="LPL_FL002_PL1"
/lab_host="DH10B"
/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 222 a 221 c 217 g 290 t 17 others
ORIGIN

Alignment Scores:
Pred. No.: 6.28e-54 Length: 967
Score: 916.50 Matches: 174
Percent Similarity: 91.15% Conservative: 1
Best Local Similarity: 90.62% Mismatches: 14
Query Match: 30.87% Indels: 5
DB: 9 Gaps: 3

US-09-763-712A-2 (1-547) x AL568743 (1-967)

QY 356 GlyThrValGlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyVal 375
Db 947 GGCACCGTGGAGCGCTGGGTCT---GGACCTCGGAGCTGCAG---CTTGCCCTGGGTA 894

QY 376 ProGlyMetProGlyProGlyProGlyProGlyProGlyProSerGlyAlaValVal 395
Db 893 CCAGCGATGCCAGSCCAGG---CCCCCCCCCTCTCTGC---CCATCAGGAGCGGTTGTG 838

QY 396 ProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProProHis 415
Db 837 CCCCCTGSCCTGCAGCAATGAGCAACCCCGGACCGGAGGACAATGGCTGCCCGCTCAC 778

QY 416 TrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAsp 435
Db 777 TGGAAAGACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTTCAGGAT 718

QY 436 AlaLysLeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlu 455
Db 717 GCAAAGCTTTTCTGTAGACAAAGTCTTCACATCTGTTTTCATAAACAAGAGAGGAA 658

QY 456 GlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAsp 475
Db 657 CAGCAATGGATAAAAAACAGATGCTAGGAGAGAGAGCCACTGGATCGCCCTCACAGAC 598

QY 476 SerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrp 495
Db 597 TCAGAGCGTGAATAATGAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATGG 538

QY 496 LysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGly 515
Db 537 AAAGCTGGACACCCGGATAACTGGGGTCTATGC-CATGGGCCAGGAGAGACTGTCTCTGGG 479

QY 516 LeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCys 535
Db 478 TTGATTATGCTGGCAGTGGAGCAATTTCCCAATGTGAAGACGTCAATAACTTCATTTC 419

QY 536 GluLysAspArgGluThrValLeuSerSerAlaLeu 547
Db 418 GAAAAAGACAGGAGACAGACTGTCTATCTGCATTA 383

RESULT 11
BQ127513
LOCUS BQ127513
DEFINITION 1160h06.y1 Melton Normalized Mixed Mouse Pancreas 1 NI-MMS1 Mus musculus cDNA clone IMAGE:5946107 5' similar to TR:Q9UM15 Q9UM15 CSR2. ;, mRNA sequence.
ACCESSION BQ127513
VERSION BQ127513.1 GI:20201424
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 581)
AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarisvili, R., Williams, T., Jackson, Y., and Bowers, Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Other_ESTs: 1160h06.x1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center This clone is available royalty-free through LBNL; please contact the IMAGE consortium (info@image.llnl.gov) for further information
Seq primer: -40RP from Gibco
High quality sequence stop: 446.
FEATURES
source
Location/Qualifiers
1..581
/organism="Mus musculus"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:5946107"
/clone_lib="Melton Normalized Mixed Mouse Pancreas 1 NI-MMS1"
/sex="Both for embryonic & newborn, male for adult and adult islet"
/dev_stages="Embryonic day 10.5, E12.5, E16.5, newborn, adult, mixed"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; Five libraries representing E10.5/12.5 pancreatic bud, E16.5 pancreas, newborn pancreas, adult pancreas, and adult islets of Langerhans were separately constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA was made by oligo-dT priming and size-selected by column fractionation. Libraries were amplified once on solid support and plasmid DNA from each library was prepared and mixed in equal amounts. The mixed library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded mixed library plasmid DNA was mixed with 5 micrograms PCR product representing mixed library inserts and hybridized to an EcoT of 6. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite

chromatography and used to make this library."

```
BASE COUNT      189 a 159 c 114 g 119 t
ORIGIN

Alignment Scores:
Pred. No.:      4,37e-53      Length:      581
Score:          901.00      Matches:      177
Percent Similarity: 97.41%      Conservative: 11
Best Local Similarity: 91.71%      Mismatches: 5
Query Match:      30.35%      Indels:      0
DB:              14      Gaps:      0

US-09-763-712A-2 (1-547) x B0127513 (1-581)

QY 12 AsnAsnLeuAsnLeuThrGlnValGlnGlnArgAsnLeuLeuThrAsnLeuGlnArgSer 31
Db 3 AACAACCTGTACCTACCCAGGTTTCAGCAGAGGAACCTTATCTCAAATCTGCAGCAGTCT 62

QY 32 ValAspAspThrSerGlnAlaIleGlnArgIleLysAsnAspPheGlnAsnLeuGlnGln 51
Db 63 GTGGATGACAAAGCCTGGCCATCCAGCGAATTAAGAATGATTTCAAAATCTGCAGCAG 122

QY 52 ValPheLeuGlnAlaLysLysAspThrAspTrpLeuLysGluLysValGlnSerLeuGln 71
Db 123 GTTTCCTTCAGCCCAAGAGGACACCGATTGGCTAAAGAAAAGTACAGAGCTTGCAG 182

QY 72 ThrLeuAlaAlaAsnSerAlaLeuAlaLysAlaAsnAspThrLeuGluAspMet 91
Db 183 ACATTGGCTGCCAACAACTCTGCCCTGGCCAAAGCAACAAATGACACCCCTAGAGGATATG 242

QY 92 AsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGlnAla 111
Db 243 AATAGCAGCTCAGCTCAATTCACAGGTCAGATGGACAACATTAACCACTATCTCACAGGCC 302

QY 112 AsnGluGlnAsnLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThrAla 131
Db 303 ACGAGCAGAGGCTGNAAGACCTTCAGGACTTACACAGGATACAGAAATAGAACAGCT 362

QY 132 IleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsnIle 151
Db 363 GTCAAGTTCAGCCAACTTGAGAACGCTTCCAGGCTTTTGACACAGATATCTGAAACATC 422

QY 152 IleSerAsnIleSerThrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsnGlu 171
Db 423 ATTAGCAACATCAGTACACAGCCCATCACCTGAGGACACTGACCAGCAATCTGAAATGAT 482

QY 172 ValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeuAsn 191
Db 483 GTTAGGACCACATGCACAGACACCTTTGACCAGACACACGGATGACCTTGACCTTGAAT 542

QY 192 AsnThrLeuAlaAsnIleArgLeuAspSerValSerLeu 204
Db 543 AACACACTAGTCAACATCGCTTGGATTCTATTCTCTC 581

RESULT 12
AW958053
LOCUS          AW958053          552 bp      mRNA      linear      EST 01-JUN-2000
DEFINITION    EST370123 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION     AW958053
VERSION       AW958053.1 GI:8147736
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 552)
AUTHORS       Heyde,P., Qi,R., Abernathy,K., Dharap,S., Gaspar,R., Gay,C., Holt
               ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
               Quackenbush,J.
TITLE         Assessment of gene expression patterns in a model of colon tumor
               metastasis using a 19,200 element cDNA microarray
JOURNAL       Unpublished (2000)
COMMENT       Contact: John Quackenbush
```

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528

Fax: 301 838 0208

Email: johnq@tigr.org

Plate: 115

Seq primer: Reverse.

Location/Qualifiers

source

1..552

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="MAGE resequences, MAGE"

/note="Vector: pBluescriptSKm"

176 a 112 c 143 g 121 t

Alignment Scores:

Pred. No.: 4,95e-49 Length: 552

Score: 842.00 Matches: 149

Percent Similarity: 98.68% Conservative: 1

Best Local Similarity: 98.03% Mismatches: 1

Query Match: 28.36% Indels: 1

DB: 10 Gaps: 0

US-09-763-712A-2 (1-547) x AW958053 (1-552)

QY 397 LeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProHisTrp 416

Db 1 CTGGCCCTGCAGCAATGAGCCCAACCCGCGCAGGAGCAATGGCTGCCCGCTCACTGG 60

QY 417 LysAsnPheThrAspLysCysTrpTyrPheSerValGluLysGluIlePheGluAspAla 436

Db 61 AAGAAGCTTCACAGACAAATGCTACTATTTTTCAGTTTGAGAAAGAAATTTTGGAGTGCA 120

QY 437 LysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGln 456

Db 121 AAGCTTTTCTGTGAGACAAAGTCTTCACATCTTGTTTTCATAAACACTAGAGAGGAACAG 180

QY 457 GlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSer 476

Db 181 CAATGGATAAAAAACAGATGCTAGGGAGAGAGAGCCACTGGCTCGGCCTCAGACTCA 240

QY 477 GluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTrpLysAsnTrpLys 496

Db 241 GAGCGTGAATGAATGGAAGTGGCTGGATGGACATCTCCAGACTACAAAAATTGGAAA 300

QY 497 AlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeu 516

Db 301 GCTGACAGCCCGGTAACCTGGGGTTCATGGCCATGGGCCAGGAGAGACTGTGCTGGTTG 360

QY 517 IleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPhe-IleCysGln 536

Db 361 ATTTATGCTGGCAGGGGAACGATTTCCAATCTGAAGACGCTCAATAACTTAATTGGCA 420

QY 536 uLysAspArgGluThrValLeuSerSerAlaLeu 547

Db 421 AAAAGACAGGGAGACAGTACTGGCATCTGCATTA 454

RESULT 13

BB248064

LOCUS BB248064 638 bp mRNA linear EST 23-OCT-2001

DEFINITION BB248064 RIKEN full-length enriched, 7 days neonate cerebellum Mus

ACCESSION BB248064

VERSION BB248064.2 GI:16355610

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 638)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,

Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda

M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M., and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
On Jul 6, 2000 this sequence version replaced gi:8940810.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

BASE COUNT 212 a 134 c 146 g 146 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3 31e-48 Length: 638
 Score: 831.00 Matches: 142
 Percent Similarity: 93.25% Conservative: 10
 Best Local Similarity: 87.12% Mismatches: 11
 Query Match: 27.99% Indels: 0
 DB: 10 Gaps: 0
 US-09-763-712A-2 (1-547) x BB248064 (1-638)
 QY 385 ProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThr 404
 |||||.....:|||||
 Db 3 CCTGCCCTCAGGCCCTCAGGAGCAATGGAGCAATTGGCTCTGCAGATGAACCAACC 62
 QY 405 ProAlaProGluAspAsnGlyCysProHisTrpLysAsnPhetHrAspLysCysTyr 424
 |||||.....:|||||
 Db 63 CCAGCATCAGAGGTCACGGATGTCGCCCTCACTGGAGAACTTCACAGATAAATGCTAC 122
 QY 425 TyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSer 444
 |||||.....:|||||
 Db 123 TATTTTCATGTGGAAAAGAAATTTTTCAGATGCTAAGCTTTTCTGTGAAGACAAATCT 182
 QY 445 SerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetVal 464
 |||||.....:|||||
 Db 183 TCCCATCTCGTTTCATAAACTCAAGAGAAAGACAGCAATGGATAAAAAAGCATACCGTG 242
 QY 465 GlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrp 484
 |||||.....:|||||
 Db 243 GGGAGAGAAAGCCATTGGATCGGCCTCACAGACTCAGACAGAGAAAGCAATGGAAAGTG 302
 QY 485 LeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGly 504
 |||||.....:|||||
 Db 303 CTAGACGGGTACCTGTTGATTACAAAACCTGGAAAGCTGGCAACACAGATAACTGGGGC 362
 QY 505 HisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAsp 524
 |||||.....:|||||
 Db 363 AGTGCCATGGCCAGCAGAGAACTGTGCTGGCTTGATTTACGCAGGACAGTGGAAATGAC 422
 QY 525 PheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSer 544
 |||||.....:|||||
 Db 423 TTCAGTGTGATGAAATCAATAACTTCAATTTGTGAGAAAGGAGGAGGCAGTACCATCA 482
 QY 545 SerAlaLeu 547
 |||||
 Db 483 TCCATATTA 491
 RESULT 14
 BF805750/c
 LOCUS
 DEFINITION QV1-C10173-071100-464-c02 C10173 Homo sapiens cDNA, mRNA sequence. EST 12-JAN-2001
 ACCESSION BF805750
 VERSION BF805750.1 GI:12134739
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 511)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV1&t2=QV1-C10173-
071100-464-c02&t3=2000-11-07&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 492.

FEATURES

source
1..511
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="C10173"
/dev_stage="Adult"
/note="Organ: colon_ins; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 94 a 113 c 139 g 165 t
ORIGIN

Alignment Scores:

Pred. No.: 1,31e-44 Length: 511
Score: 777.50 Matches: 160
Percent Similarity: 98.77% Conservative: 0
Best Local Similarity: 98.77% Mismatches: 1
Query Match: 26.19% Indels: 2
DB: 12 Gaps: 1

US-09-763-712a-2 (1-547) x BF805750 (1-511)

QY 1 MetTyrSerHisAsnValValIleMetAsnLeuAsnAsnLeuAsnLeuThrGlnValGln 20
|||||
Db 483 ATGTATTCATTAATGTGTCATCATGACCTCAACACCTGAACCTGACCCAGGTGCAG 424
QY 21 GlnArgAsnLeuIleThrAsnLeuGlnArgSerValAspThrSerGlnAlaIleGln 40
|||||
Db 423 CAGAGGAACCTCATCAGCAATCTGCAGCGGTCTGTGGATGACACACAGCCAGGCTATCCAG 364
QY 41 ArgIleLysAsnAspPheGlnAsnLeuGlnValPheLeuGlnAlaLysLysAspThr 60
|||||
Db 363 CGAATCAAGACGACTTTCAAATCTGCAGCAGGTTTCTTCAAGCCCAAGAGGACAGC 304
QY 61 AspTrpLeuLysGluLysValGlnSerLeuGlnThrLeuAlaAlaAsnAsnSerAlaLeu 80
|||||
Db 303 GATTGGCTCAAGGAGAAAGTCGACAGCTTGCAGACGCTGGCTGCCAACAACTCTGCGTTG 244
QY 81 AlaLysAlaAsnAsnAspThrLeuGluAspMetAsnSerGlnLeuAsnSerPheThrGly 100
|||||
Db 243 GCCAAGGCCAACACGACACCTTGGAGGATATGACAGCCGAGCTCAACTCATTCACAGT 184
QY 101 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 120
|||||
Db 183 CAGATGGAGAAATCATCCACTATCTCTCAAGCCACGACGACGACCACTGAAACCTGCAG 124
QY 121 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 140
|||||
Db 123 GACTTACACAAAGATGCAGAGATAGAACAGCATCAAGTTCACCAACTGGGAGAACGC 64
QY 141 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerThrAlaHis 160
|||||
Db 63 TTCCAGCTCTTTGAGACGGATATGTGAACATCATTAGC---ATACG-TACACAGCCAC 8
QY 161 HisLeu 162
|||||
Db 7 CACCTA 2

RESULT 15

BQ771366
LOCUS
DEFINITION
UI-M-F10-Dyu-g-09-0-UI.r1 NIH_BMAP_F10 Mus musculus CDNA clone
IMAGE:5702432 5', mRNA sequence.
BQ771366
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 808)
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
JOURNAL
COMMENT
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgpbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.

FEATURES

Location/Qualifiers
1..808
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5702432"
/clone_lib="NIH_BMAP_F10"
/tissue_type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according
Donaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CAGCCACGAC. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

BASE COUNT 270 a 158 c 173 g 206 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 1.11e-43 Length: 808
Score: 767.00 Matches: 131
Percent Similarity: 93.33% Conservative: 9
Best Local Similarity: 87.33% Mismatches: 10
Query Match: 25.83% Indels: 0
DB: 14 Gaps: 0

US-09-763-712a-2 (1-547) x BQ771366 (1-808)

QY 398 AlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProHisTrpLys 417
|||||
Db 2 GCTGTGCAGAAATGAACCAACCCAGCATCAGAGGTCAACGGATGTCGCGCTCACTGGAAG 61
QY 418 AsnPheThrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLys 437
|||||

```
Db 62 AACTTCACAGATAAATGCTACTATTTCATTGGAAAAAGAAATTTTGAAGATGCTAAG 121
Qy 438 LeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArgGluGluGlnGln 457
    |||||
Db 122 CTTTTCGTGAGACAAATCTCCCATCTCGTTTTTCATAAACTCAAGAGAGAACAGCAA 181
    |||||
Qy 458 TrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGlu 477
    |||||
Db 182 TGGATAAAAAAGCATACCGTGGGAGAGAAAGCCATTGGATCGGCCTCACAGACTCAGAA 241
    |||||
Qy 478 ArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAla 497
    ::|||
Db 242 CAGGAAGCGAATGGAAGTGGCTAGACGGTCACCTGTTGATTACAAAACTGGAAAGCT 301
    |||||
Qy 498 GlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIle 517
    |||||
Db 302 GGACAACCCAGATAACTGGGGCAGTGGCCATGGCCAGAGAGACTGTCTGGCTTGATT 361
    |||||
Qy 518 TyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLys 537
    |||||
Db 362 TACGAGGACAGTGGGAATGACTTCCAGTGTGATGAATCAATAACTTCATTGTGAGAAG 421
    |||||
Qy 538 AspArgGluThrValLeuSerSerAlaLeu 547
    ::|||
Db 422 GAAGGGAGGAGCAGTACCATCATCATATTA 451
    |||||
```

Search completed: March 21, 2003, 12:20:20
Job time : 1823.03 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 06:54:47 ; Search time 2261.11 Seconds
(without alignments)
4401.885 Million cell updates/sec

Title: US-09-763-712a-2_copy_206_547

Perfect score: 1940

Sequence: 1 MQDDLMSRLDETVANLSVI.....EDVNNFICKDRETVLSSAL 342

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cn2.1/USPTO.scoil/US09763712/runat_14032003.100949.18101/app_query.fasta_1.1877
-DB=GenEmbl -QFT=fastap -SUFFIX=rge -MINMATCH=0.1 -DOOPCL=0 -DOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712.@CGN.1.1.6828.0runat_14032003.100949.18101 -NCPU=3
-NO_XLPXY -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl :
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rod.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|------------|--------------------|
| 1 | 1940 | 100.0 | 2983 | 9 | AB005145 | AB005145 Homo sapi |
| 2 | 1934 | 99.7 | 2005 | 6 | AX454442 | AX454442 Sequence |
| 3 | 1934 | 99.7 | 2005 | 6 | AX490920 | AX490920 Sequence |
| 4 | 1934 | 99.7 | 2641 | 6 | AX047353 | AX047353 Sequence |
| 5 | 1934 | 99.7 | 3058 | 9 | AB038518 | AB038518 Homo sapi |
| 6 | 1932 | 99.6 | 1886 | 9 | HS802985 | AL713657 Homo sapi |
| 7 | 1801 | 92.8 | 2637 | 10 | AB078434 | AB078434 Mus muscu |
| 8 | 1795 | 92.5 | 3291 | 10 | AB038519 | AB038519 Mus muscu |
| 9 | 1664 | 85.8 | 4330 | 9 | AB052103 | AB052103 Homo sapi |
| 10 | 1021.5 | 52.7 | 130763 | 2 | AC112416 | AC112416 Rattus no |
| 11 | 981 | 50.6 | 182029 | 2 | AC114677 | AC114677 Mus muscu |
| 12 | 981 | 50.6 | 193208 | 2 | AC102618 | AC102618 Mus muscu |
| 13 | 954.5 | 49.2 | 169088 | 2 | AC016128 | AC016128 Homo sapi |
| 14 | 954.5 | 49.2 | 188439 | 9 | AP000915 | AP000915 Homo sapi |
| 15 | 935.5 | 48.2 | 71044 | 2 | AC024368 | AC024368 Homo sapi |
| 16 | 916.5 | 47.2 | 187635 | 2 | AP001022 | AP001022 Homo sapi |
| 17 | 914.5 | 47.1 | 178022 | 2 | AP000900 | AP000900 Homo sapi |
| 18 | 842 | 43.4 | 188255 | 2 | AP000939 | AP000939 Homo sapi |
| 19 | 551 | 28.4 | 2215 | 10 | BC026446 | BC026446 Mus muscu |
| 20 | 540 | 27.8 | 3636 | 9 | AB007829 | AB007829 Homo sapi |
| 21 | 540 | 27.8 | 3685 | 6 | E32511 | E32511 Scavenger r |
| 22 | 540 | 27.8 | 3810 | 6 | E32509 | E32509 Scavenger r |
| 23 | 491.5 | 25.3 | 1265 | 10 | RATSPD | M81231 Rat palmona |
| 24 | 483.5 | 24.9 | 3508 | 3 | AF053338 | AF053338 Alvinella |
| 25 | 481.5 | 24.8 | 1183 | 10 | BC003705 | BC003705 Mus muscu |
| 26 | 481.5 | 24.8 | 1253 | 10 | MUSSPD | L40156 Mus musculu |
| 27 | 474.5 | 24.5 | 1410 | 6 | AX334792 | AX334792 Sequence |
| 28 | 474.5 | 24.5 | 1410 | 9 | HSNRNAPD | X65018 H.sapiens m |
| 29 | 474 | 24.4 | 1454 | 4 | BTLSPD | X75911 B.taurus mR |
| 30 | 473.5 | 24.4 | 1301 | 9 | BC022318 | BC022318 Homo sapi |
| 31 | 473 | 24.4 | 1284 | 4 | AF509590 | AF509590 Bos tauru |
| 32 | 471 | 24.3 | 2377 | 9 | BC008760 | BC008760 Homo sapi |
| 33 | 471 | 24.3 | 5676 | 6 | E07265 | E07265 cDNA encodi |
| 34 | 471 | 24.3 | 5676 | 9 | HUMCALV | D90279 Human mRNA |
| 35 | 471 | 24.3 | 7138 | 9 | HUMPALV | M76729 Human pro-a |
| 36 | 465 | 24.0 | 6114 | 10 | CRUPALV | M76730 Chinese ham |
| 37 | 462.5 | 23.8 | 4990 | 10 | CRUCOLVITA | L08863 Cricetus |
| 38 | 462 | 23.8 | 2777 | 5 | GG007973 | U07973 Gallus gall |
| 39 | 462 | 23.8 | 4428 | 6 | AX146422 | AX146422 Sequence |
| 40 | 462 | 23.8 | 4428 | 6 | AX146424 | AX146424 Sequence |
| 41 | 462 | 23.8 | 5575 | 5 | AF137273 | AF137273 Gallus ga |
| 42 | 461 | 23.8 | 5551 | 10 | AF274662 | AF274662 Rattus no |
| 43 | 461 | 23.8 | 6156 | 10 | AB009993 | AB009993 Mus muscu |
| 44 | 459.5 | 23.7 | 810 | 6 | AR014116 | AR014116 Sequence |
| 45 | 459 | 23.7 | 5650 | 5 | AB015440 | AB015440 Rana cate |

ALIGNMENTS

RESULT 1

AB005145
LOCUS Homo sapiens CL-P1 mRNA 2983 bp mRNA linear PRI 21-NOV-2001
DEFINITION Homo sapiens CL-P1 mRNA for collectin placenta 1, complete cds.
ACCESSION AB005145
VERSION AB005145.1 GI:17026100
KEYWORDS
SOURCE Homo sapiens female tissue_lib:placenta cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H., Sakai,Y., Fukuchi,A., Sakamoto,T., Itabe,H., Suzutani,T., Ogasawara,M., Yoshida,I. and Wakamiya.N.
The membrane-type collectin CL-P1 is a scavenger receptor on vascular endothelial cells
J. Biol. Chem. 276 (47), 44222-44228 (2001)
21570232
MEDLINE 2 (bases 1 to 2983)
REFERENCE Ohtani,K.
AUTHORS Direct Submission
JOURNAL Submitted (25-JUN-1997) Katsuki Ohtani, Asahikawa Medical College,
Department of Microbiology; 2-1-1-1 Midorigaoka-Higashi, Asahikawa,
Hokkaido 078-8510, Japan (E-mail: ohtanieasahikawa-med.ac.jp,
Tel:+81-166-68-2393, Fax:+81-166-68-2399)
FEATURES
Source Location/Qualifiers
1..2983
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="female"
/tissue_lib="placenta"
1..2983
/gene="CL-P1"
71..2299
/codon_start=1
/product="collectin placenta 1"
/protein_id="BAB72147.1"
/db_xref="GI:17026101"
/translation="MKDDFAEEEEVQSFYKRGIEQGTCTCKKNWALKFESIILY
ILCALLTIVAILGKVKVMDNVGTGMETSQTQYDDKLTAVESDLKLGDQTKKAI
SNSELSFRSDILDLROQLREITEKTSKNDKTLKLOASGDALVDROSOLKTELNN
SFLITVVKTLQANGYVYNLQODTSVLQGNLQNMYSHNVIMNLNLTLTQVQRN
LITNQRSDDTSQAIRIKDFQNLQVFLQAKKDTDLWKEKVSQSLQTLAANNALAE
KANNDTLEMDNSLANSFTGMENITTI SOANEONLKDQLDLKHDAENRTAIFNLEE
RQLPFDITVNTISNISTAHRLRTLNLNEVRTCTDTLTAKHDDTLTSLNKTLANI
RLDSVLSRMOQDLMSRLDTEVANLSVIMEEMKLVDSKHGLIKNFTILOQPPGPRGP
RGDSGQPPGPGKNGKQGEKGEPPGPGPAGERGP IGPAGPPGERGKSGKSGQPK
GSRGSPGPGQPSGDPGPPGPKGEGLPGPQPGPGQGLGTGVEGPGVPGPLPG
LPGVGMFGPKGPPGPGSGAVFLALQNEPTPAPDNCGCPHWNKFNFTDKCYFSVE
KEIFEDAKLFCEKSSHLVFINTRREEQWIKQMVGRSHWTLGLTDSRENEMKWLGD
TSPDYKNMKAGQPDNMGHGPGEPCAGLIYAGQWNDFQEDVNNFICEKDRVTLSS
AL"

BASE COUNT 914 a 707 c 703 g 659 t
ORIGIN

Alignment Scores:
Pred. No.: 1,95e-58 Length: 2983
Score: 1940.00 Matches: 342
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-763-712A-2_COPY_206_547 (1-342) x AB005145 (1-2983)

QY 1 MetGlnGluAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
|||||
Db 1271 ATGCAACAAGATTGATGAGTTCAGGCTAGACACTGAAGTAGCCTATATCAGTGATT 1330
|||||
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
|||||
Db 1331 ATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTGCATCATCAAGAAATTTTACAATA 1390
|||||

QY 41 LeuGlnGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 60
|||||
Db 1391 CTACAAGGTCCACCGGCGCCCGAGGGTCCCAAGAGGTGACAGAGGATCCCGAGGACCCCT 1450
|||||
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 80
|||||
Db 1451 GGCCCAACTGGCAACAAGGACACAAAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGG 1510
|||||
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyProGlyProGlyProGly 100
|||||
Db 1511 GCGGGTGAGAGAGGCCCAATTGGACCAGCTGGTCCCGGAGAGGCTGGCGGCAAGGA 1570
|||||
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
|||||
Db 1571 TCTAAAGGCTCCACGAGGCCCAAGGCTCCCGTGGTTCCTCGTGGGAAGCCCGGCCCTCAG 1630
|||||
QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyProGlyProGlyProGly 140
|||||
Db 1631 GGCCCCAGTGGGAGCCACGAGGCCCGCGGCCACACAGGCAAGAGGAGGACTCCCGGCCCT 1690
|||||
QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
|||||
Db 1691 CAGGGCCCTCTGGCTTCAGGGACTTCAGGGCACCCTGGTGGGAGGCTGGGGTGCCTGGA 1750
|||||
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
|||||
Db 1751 CCTCGGGGACTGCCAGCTTGCCTGGGTACCAGGCATGCCAGGCCCAAGGCGCCCTCC 1810
|||||
QY 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
|||||
Db 1811 GGCCCTCTGGGCCATCAGGAGCGGTGGTGCCTGGCCCTGCAGAAATGAGCAACCCCG 1870
|||||
QY 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAspLysCysTrpTyr 220
|||||
Db 1871 GCACCGGAGGACAAATGGCTGCCCGCCCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 1930
|||||
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
|||||
Db 1931 TTTTTCAGTTGAGAAAGAAATTTTGGAGGTGCAAGCTTTCTGTGTAAGACAAAGCTTCA 1990
|||||
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
|||||
Db 1991 CATCTTGTGTTTCAATAACACTAGAGAGCAACCAATGGATAAAAAACAGATGGTAGGG 2050
|||||
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
|||||
Db 2051 AGAGAGAGCCACTGGATCGCCTCACAGACTCAGAGCGTGAAATGAATGGAAGTGGCTG 2110
|||||
QY 281 AspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
|||||
Db 2111 GATGGGACATCTCCAGACTACAAAAATTGGAAAGCTGGACAGCCGGATAACTGGGGTCAT 2170
|||||
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
|||||
Db 2171 GGCCATGGGCCAGGAGAAAGACTGTGCTGGTGTGATTATGCTGGGAGTGGAAAGATTTC 2230
|||||
QY 321 GlnCysGluAspValAsnAsnPheIleCysGlyLysAspArgGluThrValLeuSerSer 340
|||||
Db 2231 CAATGTGAAGAGCTCAATCACTTCAATTTGCGAAAAAGACAGGGAGACAGTACTGTCTCT 2290
|||||
QY 341 AlaLeu 342
|||||
Db 2291 GCATTA 2296
|||||
RESULT 2
LOCUS AX454442
DEFINITION Sequence 27 from Patent WO0208284.
ACCESSION AX454442
VERSION AX454442.1 GI:21713845
KEYWORDS human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Baker, K.P., Ferrara, N., Gerber, H., Gerttsen, M.E., Goddard, A.,
Godowski, P.J., Gurney, A.L., Hillan, K.J., Marsters, S.A., Pan, J.,
Paoni, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I.
and Ye, W.
TITLE Compositions and methods for the diagnosis and treatment of
disorders involving angiogenesis
JOURNAL Patent: WO 0208284-A 27 31-JAN-2002;
Genentech, Inc. (US) ; Baker, Kevin P. (US) ; Ferrara, Napoleone
(US) ; Gerber, Hanspeter (US) ; Gerttsen, Mary E. (US) ; Goddard,
Audrey (US) ; Godowski, Paul J. (US) ; Gurney, Austin L. (US) ;
Hillan, Kenneth J. (US) ; Marsters, Scot A. (US) ; Pan, James (US)
; Paoni, Nicholas F. (US) ; Stephan, Jean-Philippe F. (US) ;
Watanabe, Colin K. (US) ; Williams, P. Mickey (US) ; Wood, William
I. (US)
FEATURES Location/Qualifiers
source 1..2005
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 606 a 493 c 491 g 415 t
ORIGIN
Alignment Scores:
Pred. No.: 2,38e-58 Length: 2005
Score: 1934.00 Matches: 341
Percent Similarity: 99.71% Conservatives: 0
Best Local Similarity: 99.71% Mismatches: 1
Query Match: 99.69% Indels: 0
DB: 6 Gaps: 0
US-09-763-712a-2_COPY_206_547 (1-342) x AX454442 (1-2005)
QY 1 MetGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 478 ATGCAACAAGATTGATGAGTTCGAGGTGTAGACACTGAAGTAACTATCATGATT 537
QY 21 MetGluLeuMetLysLeuValAspSerLysHisGlyGlnLeuLeuLysAsnPheThrIle 40
Db 538 ATGGAGAATAAGCTAGTACCTCAAGCATGGTGCAGTCTATCAAGATTTTACAATA 597
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
Db 598 CTACAAAGTCCACCGGCCCCAGGGTCCAGAGGTGCAGAGGATCCCGAGGACCCCT 657
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGlyLysGlyGlyProGlyProGlyPro 80
Db 658 GSCCAACCTGGCAACAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProAlaGlyProGlyGluArgGlyGly 100
Db 718 GCGGTGAGAGAGGCCCAATTGACAGCTGTGTCCTCCCGGAGAGCGTGGCGCAAGA 777
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGln 120
Db 778 TCTAAAGGCTCCAGGGCCCCAAGGCTCCCGTGTTCCTCCCTGGGAAGCCCGCCCTCAG 837
QY 121 GlyProSerGlyAspProGlyProProGlyProProGlyLysGlyGlyLeuProGly 140
Db 838 GSCCCCAAGTGGGAGCCAGGCCCCCGGGCCACCAGGCAAGAGAGAGAGAGAGAGAG 897
QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
Db 898 CAGGGCCCTCCCTGCTCCAGGAGCTTCCAGGACCGCTGGGAGAGAGAGAGAGAGAG 957
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 958 CCTGGGGAGTCCAGGCTTGGCTGGGTACCCAGGATCCAGGCCCCCAAGGCCCCCCC 1017
QY 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProPro 200
Db 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProPro 200

Db 1018 GGCCCTCTCGGCCCATCAGAGGGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCCG 1077
QY 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAspLysCysTrpTyr 220
Db 1078 GCACCGGAGGACAATAGCTGCCCGCTCACTGGAAGAACTTCACACACAAATGCTACTAT 1137
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSer 240
Db 1138 TTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAAAGCTTTTCTGTGAAGCAAGTCTTCA 1197
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1198 CATCTGTGTTTTCATAAACACTAGAGAGAACACGCAATGGATAAAAAACAGATGGTAGG 1257
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 1258 AGAGAGAGCCACTGGATCGGCCTCAGACTCAGAGCGTGAAGTGAAGTGAAGTGGCTG 1317
QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 1318 GATGGACATCTCCAGACTACAAAAATTGGAAGCTGGACAGCGGATAACTGGGGTCAT 1377
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 1378 GGCCATGGGCCAGAGAGAACTGTCTGGTTCATTTATGCTGGGAGTGGACGATTC 1437
QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
Db 1438 CAATGTGAAGAGCTCANTACTTTCATTCGAAAAAGAGAGGAGAGACAGTACTGTCATCT 1497
QY 341 AlaLeu 342
Db 1498 GCATTA 1503
RESULT 3
AX490920
LOCUS AX490920 Sequence 27 from Patent WO200690.
DEFINITION AX490920
ACCESSION AX490920
VERSION AX490920.1 GI:22323797
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Baker, K.P., Ferrara, N., Gerber, H., Gerttsen, M.E., Goddard, A.,
Godowski, P.J., Gurney, A.L., Hillan, K.J., Marsters, S.A., Pan, J.,
Paoni, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I.
and Ye, W.
TITLE Compositions and methods for the diagnosis and treatment of
disorders involving angiogenesis
JOURNAL Patent: WO 0200690-A 27 03-JAN-2002;
Genentech, Inc. (US)
FEATURES Location/Qualifiers
source 1..2005
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 606 a 493 c 491 g 415 t
ORIGIN
Alignment Scores:
Pred. No.: 2,38e-58 Length: 2005
Score: 1934.00 Matches: 341
Percent Similarity: 99.71% Conservatives: 0
Best Local Similarity: 99.71% Mismatches: 1
Query Match: 99.69% Indels: 0
DB: 6 Gaps: 0
US-09-763-712a-2_COPY_206_547 (1-342) x AX490920 (1-2005)
QY 1 MetGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1 MetGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20

| | | | |
|----------|----------|---|----------------------------|
| Db | 478 | ATGCACACAGATTGATGAGGTGAGGTGAGACACTGAAGTAGACACTTATCAGTGATT | 537 |
| QY | 21 | MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle | 40 |
| Db | 538 | ATGGAAGAAATGAAGCTAGTAGACTCCAGCATGGTCAGCTCATCAAGAATTTTACAATA | 597 |
| QY | 41 | LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro | 60 |
| Db | 598 | CTACAAGGTCCACCGGCCCCAGGGTCCAAAGAGGTGACAGAGATCCCAAGGACCCCT | 657 |
| QY | 61 | GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyPro | 80 |
| Db | 658 | GGCCCCAACTGGCAAGGACAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG | 717 |
| QY | 81 | AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly | 100 |
| Db | 718 | GCGGGTGAGAGAGGCCCAATTGGACCAAGCTGGTCCCGGAGAGAGGTGGCGCAAGGA | 777 |
| QY | 101 | SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln | 120 |
| Db | 778 | TCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCGGAGAGCCCGCCCTCAG | 837 |
| QY | 121 | GlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyPro | 140 |
| Db | 838 | GGCCCCAGTGGGACCCAGGCCCGCCCCCGCCCAACAGAGGACTTCCCGCGCCCT | 897 |
| QY | 141 | GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly | 160 |
| Db | 898 | CAGGGCCCTCCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGAGCGCTGGGTGCCTGA | 957 |
| QY | 161 | ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro | 180 |
| Db | 958 | CCTCGGGGACTGCGCAGGCTTGCCTGGGGTACCAGGATGCGCAGGCCCAAGGCCCC | 1017 |
| QY | 181 | GlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro | 200 |
| Db | 1018 | GGCCCTCTGGCCCATCAGAGGCGGTGGTCCCTCGCCCTGCAGATGAGCCCAACCCG | 1077 |
| QY | 201 | AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyr | 220 |
| Db | 1078 | GCACCGGAGGACAATAGCTGCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTAT | 1137 |
| QY | 221 | PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSer | 240 |
| Db | 1138 | TTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAAAGCTTTCTGTGAAGACAAAGTCTTCA | 1197 |
| QY | 241 | HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysGlyGlnMetValGly | 260 |
| Db | 1198 | CATCTTGTTCATAAACACTAGAGAGGACAGCAATGGATAAAAAACAGATGGTAGGG | 1257 |
| QY | 261 | ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu | 280 |
| Db | 1258 | AGAGAGGCGACATGGATCGCCCTCACAGACTCAGACGCTGAAATCAATGGAAGTGGCTG | 1317 |
| QY | 281 | AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis | 300 |
| Db | 1318 | GATGGAGACATCTCCAGACTACAAAAATTGGAAGCTGGACAGCCGATAACTGGGTCTAT | 1377 |
| QY | 301 | GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe | 320 |
| Db | 1378 | GGCCATGGGCGCAGAGAGACTGTGCTGGGTGATTATGTGGCGAGTGGAGCAATTC | 1437 |
| QY | 321 | GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer | 340 |
| Db | 1438 | CAATGTGAAGACGTCAATAACTTCAATTCGAAAAAGACAGGGAGACACTGTGTCTATCT | 1497 |
| QY | 341 | AlaLeu | 342 |
| Db | 1498 | GCATTA | 1503 |
| RESULT | 4 | | |
| AX047353 | | | |
| LOCUS | AX047353 | 2641 bp | DNA linear PAT 15-DEC-2000 |

| | | | |
|---|---|---|-------|
| DEFINITION | Sequence 39 from Patent WO0068380. | | |
| ACCESSION | AX047353 | | |
| VERSION | AX047353.1 | | |
| KEYWORDS | GI:11876599 | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Cranial; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo. | | |
| AUTHORS | Bandman, O., Hillman, J.L., Tang, Y.T., Lal, P., Yue, H., Baughn, M.R., Lu, D.A. and Azimzai, Y. | | |
| TITLE | Extracellular matrix and adhesion-associated proteins | | |
| JOURNAL | Patent: WO 0068380-A 39 16-NOV-2000; | | |
| FEATURES | Incyte Genomics, Inc. (US) | | |
| Source | Location/Qualifiers | | |
| | 1..2641 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /note="Incyte ID No: 4586187CB1" | | |
| BASE COUNT | 823 a | 628 c | 608 g |
| ORIGIN | 582 t | | |
| Alignment Scores: | | | |
| Pred. No.: | 2,87e-58 | Length: | 2641 |
| Score: | 1934.00 | Matches: | 341 |
| Percent Similarity: | 99.71% | Conservative: | 0 |
| Best Local Similarity: | 99.71% | Mismatches: | 1 |
| Query Match: | 99.69% | Indels: | 0 |
| DB: | 6 | Gaps: | 0 |
| US-09-763-712a-2_COPY_206_547 (1-342) x AX047353 (1-2641) | | | |
| QY | 1 | MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle | 20 |
| Db | 912 | ATGCAACAGATTGATGAGGTGAGGTAGACACTGAAGTAGACACTATCAGTGATT | 971 |
| QY | 21 | MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle | 40 |
| Db | 972 | ATGGAAGAAATGAAGCTAGTAGACTCCAGCATGGTCAGCTCATCAAGAATTTTACAATA | 1031 |
| QY | 41 | LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro | 60 |
| Db | 1032 | CTACAAGGTCCACCGGCCCCAGGGTCCAAAGAGGTGACAGAGGATCCCAAGGACCCCT | 1091 |
| QY | 61 | GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyPro | 80 |
| Db | 1092 | GGCCCCAACTGGCAAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG | 1151 |
| QY | 81 | AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly | 100 |
| Db | 1152 | GCGGGTGAGAGAGGCCCAATTGGACCAAGCTGGTCCCGGAGAGAGGTGGCGCAAGGA | 1211 |
| QY | 101 | SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln | 120 |
| Db | 1212 | TCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCGGAGGACCCCGCCCTCAG | 1271 |
| QY | 121 | GlyProSerGlyAspProGlyProProGlyProProGlyLysGlyGluGlyLeuProGlyPro | 140 |
| Db | 1272 | GGCCCCAGTGGGACCCAGGCCCGCCCCCAGGCAAGAGGAGGACTCCCGCCGCT | 1331 |
| QY | 141 | GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly | 160 |
| Db | 1332 | CAGGGCCCTCCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGAGCGCTGGGTGCCTGA | 1391 |
| QY | 161 | ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro | 180 |
| Db | 1392 | CCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAGGATGCCAGGCCCAAGGCCCCCT | 1451 |
| QY | 181 | GlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro | 200 |
| Db | 1452 | GGCCCTCTGGCCCATCAGAGGCGGTGGTCCCTCGCCCTGCAGATGAGCAACCCCG | 1511 |
| QY | 201 | AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyr | 220 |

```

|||||
Db 1512 GCACCGGAGCAATAGTCCCGCTCCTCTGGAAGAACTTCACAGACAAATGCTACTAT 1571
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
Db 1572 TTTTCAGTTGAGAGAAATTTTTCAGGATGCAAGCTTTCTGTGAGACAAAGTCTTCA 1631
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1632 CATCTGTTTTCATAAACACATAGAGGAAACAGCAATGGATAAAAAACAGATGCTAGG 1691
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 1692 AGAGAGAGCCACTGATCGGCTCAGACTCAGAGCTGAAATGAATGGAAGTGGCTG 1751
QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 1752 GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGACAGCCGAGTAATCGGGTCTAT 1811
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyraLaglyGlnTrpAsnAspPhe 320
Db 1812 GGCCATGGCCAGGAGAACTGTCTGGTTGATTTATGCTGGCAGTGGAAACGATTTC 1871
QY 321 GlnCysGluAspValAsnAsnPhelCysGluLysAspArgGluThrValLeuSerSer 340
Db 1872 CAATGTGAAGAGCTCAATAACTTCTTTCGGAAGAAACAGAGGAGACAGTACTGTCTCT 1931
QY 341 AlaLeu 342
Db 1932 GCATTA 1937

RESULT 5
AB038518
LOCUS AB038518 3058 bp mRNA linear PRI 08-MAR-2001
DEFINITION Homo sapiens SRC1 mRNA for scavenger receptor with C-type lectin
type I, complete cds.
ACCESSION AB038518
VERSION AB038518.1 GI:13365514
KEYWORDS
SOURCE Homo sapiens tissue_lib:Placenta cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (sites)
Nakamura, K., Funakoshi, H., Miyamoto, K., Tokunaga, F. and Nakamura, T.
Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRC1), a novel member of a
scavenger receptor family
Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
21092718
2 (bases 1 to 3058)
Nakamura, K. and Nakamura, T.
Direct Submission
Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical
School, Division of Biochemistry, Biomedical Research Center; 2-2
Yamadoko, Suita, Osaka 565-0871, Japan
(E-mail:knakamura@onbich.med.osaka-u.ac.jp,
Tel:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)
FEATURES
Location/Qualifiers
1..3058
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_lib="Placenta"
1..3058
/gene="SRC1"
209..2437
/gene="SRC1"
/codon_start=1
/product="scavenger receptor with C-type lectin type I"
/protein_id="BAB39147.1"
/db_xref="GI:13365515"
/translation="MKDDFAEEEVQSGYKRFQIGEGTQCTCKNNWALKFSILLY
LTCALLTITVAILGKVVYKRMNVTTGGMETSRQTYDDKLTAVESDLKLGDTGKKA

```

```

STNSELSTFRSDILDROOLREITEKTSKNKDTLEKLAQSGDALVDROSLQKETLENN
SFLTTTWNKTLQAYNYVTNIQDDTSVLQNLQNMYSNNVIMNLNLNTIQOORN
LITNLRQSVDDTSOAIORIKNDFONLQOVFLQAKDQDLKKEVQSLOTLAANNSALA
KANNDLTDMNSQLNSFGOMENITTSOANEONLKDQLDHLKDAENRTAKFNQLEER
RFQLPETDVIINIISNYSYTAHHLRTLSNLNEVTTCTDTLTTKHDDTLSTNLNTANI
RLDSVSURMOODLMLRSRLDTSEANLSVIMEEMKLVDKHLGQILKNFTLIQPGPPRGP
RGDRSGQPGPTNGKQKGEKPGEPGAPGERGPIGPAGPPGERGKSGKSGQPK
GSRSPGKPGQSPGDPGPGPGKEGLPGQPGPGQGLQGTGEGVPGPGRLPG
LPVPGMGPKPGPPGPGVLAQNEPTPAEDNSCPPHWKNTDKCYFVSVE
KEIFEDAKLFCEDKSSHLVFINTRREQQWIKQMVGRSHWIGLITDSERENWKWLDG
TSPDYKNWKACQPDNWHGCHGPGEDCAGLIYAGOWNDPQCEDVNNFICEKRETVLSS
AL"
polyA_signal 3020..3025
BASE COUNT 911 a 744 c 764 g 639 t
ORIGIN
Alignment Scores:
Pred. No.: 3.17e-58 Length: 3058
Score: 1934.00 Matches: 341
Percent Similarity: 99.71% Conservative: 0
Best Local Similarity: 99.71% Mismatches: 1
Query Match: 99.69% Indels: 0
DB: 9 Gaps: 0
US-09-763-712A-2_COPY_206_547 (1-342) x AB038518 (1-3058)
QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1409 ATGCAACAAGATTGATGAGTCCAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 1468
QY 21 MetGluGluMetLysLeuValAspSeryHisGlyGlnLeuIleLysAsnPhetrIle 40
Db 1469 ATGGAAGAAATCAAGCTAGTACACTCCAAAGCATGGTCACTCATCAAGAAATTTTACAATA 1528
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
Db 1529 CTACAAGGTCCACCGGGCCCGCCAGGGGTCCAGAGGTGCACAGAGGATCCCGAGGACCCCT 1588
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
Db 1589 GGCCCACTGGGCAACAGGACAGAAAGAGAGAGAGGGGAGCCCTGGACCACTGGCCCT 1648
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly 100
Db 1649 GCGGGTGAAGAGGCCCAATTGACAGCTGGTCCCGCCGAGAGCGTGGCGCAAGGA 1708
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGln 120
Db 1709 TCTAAGGCTCCCGGGCCCGCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAG 1768
QY 121 GlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyPro 140
Db 1769 GGCCCACTGGGAGCCAGCCCGCCCGCCCGCCAGGCAAGAGAGGACTCCCGGCCCT 1828
QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
Db 1829 CAGGGCCCTCTCTGGCTTCCAGGACTTCCAGGACCGCTTGGGAGCGCTGGGTGCTCGA 1888
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 1889 CCTCGGGGACTGCCAGGCTTGCTGGGTGACAGGATGCCAGGCCCCCAAGGGCCCCCCC 1948
QY 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
Db 1949 GGCCCTCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCG 2008
QY 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPhetrAspLysCysTyTr 220
Db 2009 GCACCGGAGGACAAATAGTCCCGCTCACTGGAAGAACTTCACAGACAAATGCTACTAT 2068
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240

```

```
Db 2069 TTTTCAGTTGAGAAATAATTTTGAGGATGCAAAAGCTTTTCTGTGAGAACAGTCTTCA 2128
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 2129 CATCTTGTTTTCATAAACACTAGAGAGGAACAGCAATGATGATAAAAAACAGATGCTAGG 2188
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 2189 AGAGAGAGCACTGGATCGCCCTCACAGACTCAGAGCGTGAAATGAATGGAAGTGGCTG 2248
QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 2249 GATGGACATCTCCACACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAT 2308
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 2309 GGCCATGGCCAGGAGAACACTGTGCTGGGTGATTTATGCTGGGCGAGTGGAAAGCATTTTC 2368
QY 321 GlnCysGluAspValAsnAsnPhelIleCysGluLysAspArgGluThrValLeuSerSer 340
Db 2369 CAATGTGAAGAGCTCAATAACTTCATTTGCGAAAAAGACAGGAGACAGTACTGTCTATCT 2428
QY 341 AlaLeu 342
Db 2429 GCATTA 2434

RESULT 6
HSM802985 1886 bp mRNA linear PRI 20-MAR-2002
DEFINITION Homo sapiens mRNA; cDNA DKFZp547G1215 (from clone DKFZp547G1215).
ACCESSION AL713657
VERSION AL713657.1 GI:19584339
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Koehrer,K., Beyer,A., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE Submitted
JOURNAL Direct Submission
COMMENT Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFZp547G1215) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
FEATURES
Location/Qualifiers
1..1886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18p11.3"
/clone="DKFZp547G1215"
/tissue_type="brain"
/clone_lib="547 (synonym: hfbrl). Vector pSportli; host
DH10B; sites NotI + SalI"
/dev_stages="fetal"
1..1886
/gene="DKFZp547G1215"
<1..1256
/gene="DKFZp547G1215"
/note="similarity to asialoglycoprotein receptor (Mus
musculus"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAD28466.1"
/db_xref="GI:19584340"
/translation="AIKENQLERFQLFETDIVNIISNISTAHHLRLTSLNLEVRT
```

```
TCTDTLTAKHTDDLTSLNLTANIRLDSVSLRMQDLMRSLDTEVANLSVIMEEMKLY
DSKHGLIKNFTIILQGPFGPRGDRSQGPPTGNKQKGERGEPGPPAGERG
PIGAPAPGERGGKSGSQGPKSGRSPKPGQSGSDGDPGPPGPKGEGLPQGP
PGFOLGQTVGEPGVPGLPGLPGVPMGPKGPPGPGSGAVVPLALQNEPTPA
PEDNGCPHHKKNFTDKCYFSEVEIFEADAKLFCEDKSSHLVFINTRREQQWIKQMV
GRESHWIGLTDSERENKWLDDGTSFDYKNMKAGQPDNMGHGHGPGEDCAGLIYAGOW
NDFQCEDYNNFICEKDRFVLSAL"
1838..1843
/polya_signal /gene="DKFZp547G1215"
1869..
/polya_site /gene="DKFZp547G1215"
BASE COUNT 569 a 444 c 445 g 428 t
ORIGIN
Alignment Scores:
Pred. No.: 2,67e-58 Length: 1886
Score: 1932.00 Matches: 341
Percent Similarity: 99.71% Conservative: 0
Best Local Similarity: 99.71% Mismatches: 1
Query Match: 99.59% Indels: 0
DB: 9 Gaps: 0
US-09-763-712a-2_COPY_206_547 (1-342) x HSM802985 (1-1886)
QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 228 ATGCAACAAGATTGTGATGAGTTCGAGGTTAGACACTGAAGTAGCCAACTATCAGTGATT 287
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlnLeuLysLysAsnPhelIle 40
Db 288 ATGGAAGAATGAAGCTAGTAGACTCCAAGCATGTGTCACTCATCAAGAATTTTACAATA 347
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyProArgGlyAspArgGlySerGlnGlyPro 60
Db 348 CTACAGAGTCCACCGGGCCCCAGGGGTCCACAGAGGTGACAGAGGATCCACGAGGCCCT 407
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
Db 408 GCCCAACTGGCAACAAGGACAGAAAGAGAGAGAGGGGGGCGCTGGACCACTGGCCCT 467
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlyGluArgGlyGlyLysGly 100
Db 468 CGGGGTGAGAGAGGCCCAATTTGACACAGCTGTCTCCCGGAGAGCGTGGCGCAAGA 527
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
Db 528 TCTAAAGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTTGGGAAGCCCGCCCTCAG 587
QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyProGlyLysGluGlyLeuProGly 140
Db 588 GGCTCCAGTGGGAGCCAGGCCCCCGGGCCACCAGGCAAGAGAGGAGGACTCCCGGCCCT 647
QY 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
Db 648 CAGGGCCCTCTCGGCTTCAGGGACTTCAGGGACCGTTGGGGAGCGCTGGGTGGCTGGA 707
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 708 CCTCGGGAGTCCAGGCTTGGCTGGGTACCAAGGATCCAGGCCCCCAAGGCCCCCC 767
QY 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
Db 768 GGCCCTCTCGGCGCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCG 827
QY 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPhelIleLysLysCysTrpTyr 220
Db 828 GCACCGGAGGACAAATGGCTGGCCCTCACTGGAGAACTTCACAGACAAATGCTACTAT 887
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
Db 888 TTTTCAGTTGAGAAAGAAATTTTGGAGGATGCAAAAGCTTTTCTGTGAGACAAAGCTTCA 947
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
```

```

Db 948 CATCTGTTTTCATAAACTAGAGGACAGCAATGGATAAAAAAGACAGATGGTAGGG 1007
Qy 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluAtrGluAsnGluTrpLysTrpLeu 280
Db 1008 AGAGAGACCACTGGATGGCTCCACAGACTCAGAGCGTGAATGGAATGGAGTGGCTG 1067
Qy 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 1068 GATGGACATCCAGACTACAAAAATTGGAAGCTGGACAGCCGGAATAACTGGGTCAT 1127
Qy 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 1128 GGCCATGGCCAGGAGAGACTGCTGGTGGTGAATTTATGCTGGCAGCTGGAACGATTC 1187
Qy 321 GlnCysGluAspValAsnAsnPhelIleCysGluLysAspArgGluThrValLeuSerSer 340
Db 1188 CAATGTGAAGAGCTCAATACTTCATTTGCGAAAAAGACAGAGGAGACACTACTGTCT 1247
Qy 341 AlaLeu 342
Db 1248 GCATTA 1253

RESULT 7
AB078434 LOCUS
DEFINITION Mus musculus CL-P1 mRNA for collectin placenta 1, complete cds.
ACCESSION AB078434
VERSION AB078434.1 GI:21901968
KEYWORDS
SOURCE Mus musculus tissue_lib:Liver cdna to mRNA.
ORGANISM Mus musculus
REFERENCE
AUTHORS Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H., Sakai,Y., Fukuoka,A., Sakamoto,T., Itabe,H., Suzutani,T., Ogasawara,M., Yoshida,I. and Wakamiya,N.
TITLE cDNA cloning of mouse CL-P1 gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2637)
AUTHORS Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H., Sakai,Y., Fukuoka,A., Sakamoto,T. and Wakamiya,N.
TITLE Direct Submission
JOURNAL Submitted (18-JAN-2002) Katsuki Ohtani, Asahikawa Medical College, Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa, Hokkaido 078-8510, Japan (E-mail: ohtani.easahikawa-med.ac.jp, Tel:81-166-68-2393, Fax:81-166-68-2399)
FEATURES
source
    1..2637
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /tissue_lib="Liver"
    1..2637
        /gene="CL-P1"
        /codon_start=1
        /product="collectin placenta 1"
        /protein_id="BAC05523.1"
        /db_xref="gi:21901969"
        /translation="MKDDFAEEVEVSFGYKRFQIGTQCTCKNNWALKFSTVLVY
        STNSELSTFRSDLDLRLQLOEITEKTSKNKDLTKLEKQANGSLVDRQSOLKETLQNN
        ILCALLITVAILGYKVVEKMDNVTDGNETSHQYDNKLTAVESDLKLGDOACKAL
        SFLNTYNTKLTQANGYTNLQDTSVLQGNLQSNQMSYVIMNKNLNTLQVQRN
        LSNLQSDVDTSLAIQRIKNDLFQAKRDDIDKVKVQSLOTLANNSALA
        KANDTLEDNMSLSSFTGQMDNTTISQANEQSLKDLQDLKHDTENRTAVKFSQLE
        RFQVFETDVNIIISNYTAHLRTLSNLNMDVRTCTDTLTRHTDDTLSTNNLTVNI
        KDRGSISLRMOODMRKSLDKTEKNGEPPGAGERTGIPVGPGERSGSKSGSQGP
        KLSRSGSPGPGTCKNGKQEGEPGPGGAGRTGIPVGPGERSGSKSGSQGP
        GSRSGPKPGQSGSGGPGPGKDLPGPGPGFQGLQGTGVGFGPGPGPLPG
        LPVPGMPGPGPGPGPGMEPLALQNEPTPASEVNGCPHWNKFTDKCYFSLG
        KEIFEDAKLFCEDKSSHLVFINREEQQWIKKHTVVGRESHWIGLTDSEQSEWKLGD

```

```

SPVDYKNWKAGQPDNWSGHPGEDCAGLIYAGWDFQCDEINNFICEKREAVPSS
IL"
BASE COUNT 829 a 617 c 635 g 556 t
ORIGIN
Alignment Scores:
Pred. No.: 9.37e-54 Length: 2637
Score: 1801.00 Matches: 313
Percent Similarity: 95.91% Conservative: 15
Best Local Similarity: 91.52% Mismatches: 14
Query Match: 92.84% Indels: 0
DB: 10 Gaps: 0
US-09-763-712a-2_COPY_206_547 (1-342) x AB078434 (1-2637)
Qy 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1292 ATGCAGCAGACATGATGAGTCAAGATTAGACACTGAAGTGGCCAATTTACAGTGGTT 1351
Qy 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhThrIle 40
Db 1352 ATGGAAGAGATGAACTGGTTGACTCCCAAGCACGGTCAGCTCATCAAGAACTTTACCAT 1411
Qy 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
Db 1412 CTACAAAGTCTCTCTGGGCCAGAGGTCCTCAAGAGGTGACAGAGGATCTCAGGAGCACCT 1471
Qy 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyPro 80
Db 1472 GGTCCAACCTGGCAACAAGGACAGAAAGAGAGAGAGGAGAGGAGGCTGGTCCACTGGCCCT 1531
Qy 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyProGlyGluArgGlyGlyLysGly 100
Db 1532 GCGGCTGAGAGGGGACCAATTGACCAAGTCGCGCCCTCTCTGGAGAGCGTGGCAGCAAGCA 1591
Qy 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
Db 1592 TCCAAAGGCTCACAGGGTCCCAAGGATCTCGTGGGTCCCGGAGGAGGCTGGCCCTCAA 1651
Qy 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyPro 140
Db 1652 GGACCTAGTGGGGACCCAGGACCAAGGTCACAGGCAAGGAGGAGGAGGAGGAGGAGGAGG 1711
Qy 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
Db 1712 CAGGCGCTCTCTGGCTTCCAGGACTACAGGCACTGTGGGTGAGCTGGAGTACCTGGA 1771
Qy 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 1772 CCTCGGGGTTGCCAGGCTTGCAGGGGTGCCAGGCATGCCCTGGGCTTAAGGAGCACCT 1831
Qy 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
Db 1832 GGCCCTCCAGGCGCTCAGGAGCAATGGAGCCATTGGCTCTCGCAATGAACCAACCCCA 1891
Qy 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPhThrAspLysCysTyrTrp 220
Db 1892 GCATCAGAGGTACACGGATGTCGCTCACTGGAAGAACTTCACAGATAAATGCTACTAT 1951
Qy 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
Db 1952 TTTTTCATTGGAAAAAGAAATTTTGAAGATGTAAGCTTTTCTGTGCAAGACAAATCTTCC 2011
Qy 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 2012 CATCTCGTTTTCATAAACTCAAGAGAAACAGCAATGGATGATAAAAGCATACCTGGGG 2071
Qy 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluAtrGluAsnGluTrpLysTrpLeu 280
Db 2072 AGAGAAAGCCATTGGATCGGCTTCACAGACTCAGAACAGCAAGCAATGGAGTGGCTA 2131
Qy 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300

```

```
Db 2132 GACGGTCACTGTTGATTACAAAACTGGAAGCTGGACACACAGATAACTAGGGGCACT 2191
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2192 GGGCATGGCCAGGAGAACTGCTGCTGCTGATTACCGCAGGACAGTGGATGACTTC 2251
QY 321 GlnCysGluAspValAsnAsnPhelLeuLysGlyGlnLysAspArgGluThrValLeuSerSer 340
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2252 CAGTGTGATGAATCAATCAATTCATTTCTGTGAGAGGAAGGAGGACAGTACCATCATCC 2311
QY 341 AlaLeu 342
|||||
Db 2312 ATATTA 2317

RESULT 8
AB038519 3291 bp mRNA linear ROD 18-JAN-2002
LOCUS AB038519 Mus musculus src1 mRNA for scavenger receptor with C-type lectin,
DEFINITION complete cds.
ACCESSION AB038519
VERSION AB038519.1 GI:18146951
KEYWORDS Mus musculus cDNA to mRNA.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Nakamura, K., Funakoshi, H., Tokunaga, F. and Nakamura, T.
TITLE Molecular cloning of a mouse scavenger receptor with C-type lectin
(SRCL)(1), a novel member of the scavenger receptor family
JOURNAL Biochim. Biophys. Acta 1522 (1), 53-58 (2001)
MEDLINE 21575692
REFERENCE 2 (bases 1 to 3291)
AUTHORS Nakamura, K. and Nakamura, T.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical
School, Division of Biochemistry, Biomedical Research Center; 2-2
Yamadaoka, Suita, Osaka 565-0871, Japan
(E-mail: knakamura@onbich.med.osaka-u.ac.jp,
Tel:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)
FEATURES
Location/Qualifiers
1..3291
/organism="Mus musculus"
/db_xref="taxon:10090"
1..3291
/gene="src1"
77..2305
/gene="src1"
/codon_start=1
/product="scavenger receptor with C-type lectin"
/protein_id="BAB82497.1"
/db_xref="GI:18146952"
/translation="MKDVFAGEEEVQSFYKRFGLIHECTQCTKINNWLAKFSIVLLY
ILCALLTIVAILGVKVEKNDVSDGMETSHQYDNKLTAVEDSLKLGDOACKKAL
STNSELSTFRDLDLRQLOEITEKTSKNDTLEKLAANGDSLVDROSQKETLQNN
SFLITVNTKIQANGYVYNLQDDTNVLQNLQSQSVIMNNLNLTQVQRN
LISNLQSDSDSLAIORIKNDQFNLOQVFLQAKDTDLKQVSLQSLTAANNLSALA
KANNDLDMNSQSLSSTGQMDNITTISQANEQSLKQLDLKDKDTENRTAVKFSOLE
RFOVETDIWNIISNYSYTAHLRLTSLNLDVMTCTDTLTRHTDDLTSLNNTLVNI
RLDSILRMQDMWRSKLDTEVANLSVVMEEKLVDSKHGQLIKNFTILQPPGPRGP
KDRGSQGPPTGNKGQKGEKGPPEGPAGERGTIGPVGPGERGSKGSGQGP
GSRSPGKPGPGPGPGPGKDLPGPPGPGPGFQGLQGTVGEPVGPGRGLPG
LPVPGMPGKPGPGPGPGAMEPLALQNEPTFASVNGCPPHWKMTDRCYFSLK
KEILDALKFCEDKSHLVFINSRREQQWIKKHTVGRSHWIGLTDSEQSEMKWLDG
SPVDYKNKNAQPDNWSGSHGPGEDCAGLIIVAGQWNDPQCDIEINNFICEKREAVPSS
IL"
polyA_signal 3246..3251
/gene="src1"
BASE COUNT 1013 a 757 c 759 g 762 t
ORIGIN
Alignment Scores:
Pred. No.: 1.74e-53 Length: 3291
```

```
Score: 1795.00 Matches: 312
Percent Similarity: 95.61% Conservative: 15
Best Local Similarity: 91.23% Mismatches: 15
Query Match: 92.53% Indels: 0
Db: 10 Gaps: 0

US-09-763-712a-2_COPY_206_547 (1-342) x AB038519 (1-3291)
QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1277 ATGCACAGACATGATGAGGTCAAAGTTAGACACTGGAAGTGGCCAACTATACAGTGGT 1336
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhThrIle 40
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1337 ATGGAAGAGATCAAACTGGTTGACTCCCAAGCAGCGGTCACTCATCAAGAACTTACCAT 1396
QY 41 LeuGlnGlyProGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1397 CTACAAGGTCTCTCTGGCCCCAGAGGTCCAAAAGGTGACAGAGGATCTCAGGACCACT 1456
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1457 GGTCCAACCTGGCACACAGGACAGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1516
QY 81 AlaGlyGluArgGlyProTleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1517 GCGGGTGAGAGGGGCACAAATTTGGACCACTGCGCCCTCTCTGGAGAGCGTGGCAGCAA 1576
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGln 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1577 TCCAAGGCTCAAGGGTCCCAAGGATCTCGTGGGTCCCCAGGGAGAGCTGGCCCTCAA 1636
QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyPro 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1637 GGAACCTAGTGGGACCCAGGACCACTGCTCCACAGGCAAGATGAGTCTCCCTGGCCCT 1696
QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1697 CAGGGCCCTCTCTGGGTTCACAGGAGCTACAGGAGCTGTGGGTGAGCGTGTACCTGGA 1756
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1757 CCTCGGGGTTCAGGAGCTTGCAGGGGTGCCAGGAGTCCCTGGGCTTAAGGACCACT 1816
QY 181 GlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1817 GSCCTCCAGGCCCCCTCAGGAGCAATGGAGCCATTTGGCTCTGCAGAAATCAACCA 1876
QY 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPhThrAspLysCysTyrTyr 220
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1877 GCATCAGAGGTCAACGGATGTCGCGCTCACTGGAAGAACTTCACAGATAAATGCTACT 1936
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1937 TTTTCATGGAAAAGAAATTTTGGAAAGATGCTAAGCTTTTCTGTGAAGACAAATCTCC 1996
QY 241 HisLeuValPheIleAsnTrpArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1997 CATCTCGTTTTCATAAATCAAGAGAAAGAACAGCAATGATAAAAAAGCATACCGTGG 2056
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2057 AGAGAAAGCCATTGGATCGGCTCACAGACTCAGAGACTCAGAGAAAGCAAGCAATGGA 2116
QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2117 GACGGGTCACTGTTGATTACAAAACCTGGAAGCTGGCAACACAGATAACTTGGGCACT 2176
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2177 GGCCATGGCCAGGAGAGACTGTCTGGCTTGATTTACGAGGAGACAGTGGGAATGACTTC 2236
QY 321 GlnCysGluAspValAsnAsnPhelLecCysGluLysAspArgGluThrValLeuSerSer 340
```



```
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 2237 CAGTGTGATGAATAAATACTTCTTTGTGAGAGGAAAGGAGGAGTACCATCATCC 2296
Qy 341 AlaLeu 342
|||
Db 2297 ATATTA 2302

RESULT 9
AB052103 Homo sapiens SRC1 mRNA for scavenger receptor with C-type lectin
LOCUS AB052103 linear PRI 08-MAR-2001
DEFINITION type II, complete cds.
ACCESSION AB052103
VERSION AB052103.1 GI:13365552
KEYWORDS
SOURCE Homo sapiens cdna to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Nakamura,K., Funakoshi,H., Miyamoto,K., Tokunaga,F. and Nakamura,T.
Molecular cloning and functional characterization of a human
scavenger receptor with C-type lectin (SRC1), a novel member of a
scavenger receptor family
Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
JOURNAL 21092718
MEDLINE 2 (bases 1 to 4330)
REFERENCE Nakamura,K. and Nakamura,T.
AUTHORS Direct Submission
TITLE Submitted (04-DEC-2000) Kenji Nakamura, Osaka University Graduate
JOURNAL School of Medicine, Division of Biochemistry, Biomedical Research
Center; 2-2 Yamadaoka, suite, Osaka 565-0871, Japan
(E-mail:knakamur@onbich.med.osaka-u.ac.jp, Tel:81-6-6879-3783,
Fax:81-6-6879-3789)
FEATURES
source Location/Qualifiers
1..4330
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..4330
/gene="SRC1"
3..1871
/gene="SRCL"
/codon_start=1
/product="scavenger receptor with C-type lectin type II"
/protein_id="BAB39148.1"
/db_xref="GI:1336553"
/translacion="MKDDFAEEVEVPSFGKRFGIHGTQCPKCHNNWALKFSILLY
ILCALLTITVAIGYKVKVRMDNVGMEYSRQTYDDKLTAVESDLKGLDQTKKAI
STNSLSTFRSDIIDLRLQQLREITEKTSKNKDTLEKLAQSGDALVDRSQSLTEKNN
SFLITTVNKTQLQAYGVYTNLQDTSVLQGNLQNMYSNHYVIMNLLNLTQVQQRN
LITNLRSDVDTSOAIQRIKNDLFONLQVFLQAKKDTDLWKEKVSQSLQTLAANNSALA
KANNDTLEDNMSOLNSETGOMENITTIISOANEQNLKQLODLKDAENRTAKFNQLEE
RFQFEVDVNIINISYTAHLTLNLSNEVRITCTDITLTKHTDITLNLNLTANI
RLDSVRMOQDLMSRLDTEVANLSVIMEEMKLVDKSHGQLIKNFTILOQPPGPRGP
RDGSGQPGPTKNGKQKEGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
GSRGSPKPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
LPVPGPMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG
BASE COUNT 1186 a 1065 c 1035 g 1044 t
ORIGIN

Alignment Scores:
Pred. No.: 5,86e-49 Length: 4330
Score: 1664.00 Matches: 342
Percent Similarity: 39.13% Conservative: 0
Best Local Similarity: 39.13% Mismatches: 0
Query Match: 85.77% Indels: 532
DB: 9 Gaps: 1

US-09-763-712a-2_COPY_206_547 (1-342) x AB052103 (1-4330)

Qy 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerVal11e 20
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 1203 ATGCAACAAGATTTTGTAGGTCGAGTTAGACACTGAAGTACCAACTTATCAGTGATT 1262
```

```
Qy 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLeuLysAsnPhethrile 40
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 1263 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGTCAGCTCATCAAGAATTTTACAATA 1322
Qy 41 LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 1323 CTACAAAGGTCCACCGGGGCCCAAGAGGTCCAAAGAGGTGACAGAGGATCCCAAGGACCCCT 1382
Qy 61 GlyProThrGlyAsnLysGlnLysGlyGluLysGlyGluProGlyProGlyProGlyPro 80
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 1383 GGCCCACTGGCAACAAGGACAGAAAGAGAGAGGGGGGAGGCTGGACCACTGGCCCT 1442
Qy 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly 100
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 1443 GCGGGTGAGAGAGGCCCAATTGGACAGCTGCTCCCCCGGAGAGCGTGGCGCAAGGA 1502
Qy 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 1503 TCTAAAGGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCTTGGGAAGCCCGCCCTCAG 1562
Qy 121 GlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyPro 140
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 1563 GGCCCACTGGGGAGGCCAGGCCCGCCCGGCCCAAGAGGAGGACTCCCCGGCCCT 1622
Qy 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 1623 CAGGGCCCTCTGGCTTCCAGGGACTTCCAGGGACCGCTTGGGGAGCGCTGGGTGCTGA 1682
Qy 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 1683 CCTCGGGGACTGCCAGGCTTGGCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCGCC 1742
Qy 181 GlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 1743 GGCCCTCTGGGCCATCAGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCACCCCG 1802
Qy 201 AlaProGluAspAsn-----
|||||.....:|||||.....:|||||.....:|||||... ||| |||||
Db 1803 GCACCGGAGGACAATAGTAGTCCAAAGCCCTCCCTCCAGCGGGGAGGAGGAGTGA 1862
Qy 205 -----
Db 1863 TGTGCTTGAGACTGAGCAACCCCAAGTGTGTGGCCCAACCATGTTATTTTGTGTTTG 1922
Qy 205 -----
Db 1923 TTTTGTGTTTGTAGACAGGCTCTCACTCTGTTGCCAGGCTGGAGTGCAGTGGCGTGATCT 1982
Qy 205 -----
Db 1983 CGGCTCACTGCAACCCCAACATCCCGGGTTCAAGTGATTCTCTCGCTTAGCCTCCAGG 2042
Qy 205 -----
Db 2043 TAGCTGGGAATACAAGCGGTGTGCCCAACAGCAGCAAGTAATTTTGTATTTTAGTAGAGA 2102
Qy 205 -----
Db 2103 GGGGTTTTCACAGTGTGGGACAGGTTTTCAAACTTCTGGCCTCAAGTGATCCACCTAC 2162
Qy 205 -----
Db 2163 CTAGTTTCCCAAGTGTGGGAATACAGGTGTGAACCAACCAAGTTTGTGCCCCGGCTGC 2222
Qy 205 -----
Db 2223 ATATTATTTAACAGCAAGAGGCATCATAGGGGTTGGCACAGAGGCCCTGCTGCCAGAAGCA 2282
Qy 205 -----
Db 2283 GCCTCCATACCAGGCATGTCTCTCTCTAAAGCCATTTGGGTTTTTTTAGTCTCATTTG 2342
```

QY 205 ----- 205
Db 2343 TTTTATGATTATGTGTATGTATGATGGACAGGGTCTCACTGTCTACCCAGCGTGAAT 2402
QY 205 ----- 205
Db 2403 GACGAAGTGAAGTAGTGCAACATAGCTCATTGCGACGTTTGACCTCTCGGGCTTAGGTGA 2462
QY 205 ----- 205
Db 2463 TCTTCCACTCAGACTCTCGCTAGCTGGGACACAGGGGTATGCCACACACCTGGCTA 2522
QY 205 ----- 205
Db 2523 ATTTTATTTTCTTTTGTAGAAACAGGGTCTTGCTATGTTGCCAGGCTGGTCTTGA 2582
QY 205 ----- 205
Db 2583 ACTCTGGGCTCAAGCAACCTCTGCGCTCAGCTCTTGGCAAGTGTGGGATTCAGAGTGG 2642
QY 205 ----- 205
Db 2643 GAGCCACTGCACCTGCTCTTGCACCTGTTTATTTATAGGAGCCAGCCCTCAGAAACAA 2702
QY 205 ----- 205
Db 2703 ACACAATGCAAAACATTTCTTTAGGGAAGCCTGCTGGCGCCCTGAGGCTCTTCCCTC 2762
QY 205 ----- 205
Db 2763 GTTTTACATCTCAGCTGGAGATAGATGGGCTCCCTCGGTAGCTCCATCAACATGGG 2822
QY 205 ----- 205
Db 2823 ATGGAACCTCTGCTATTGATGACTACATTTAATTTCTTTTGGAGAAAGGTGTAGATGT 2882
QY 205 ----- 205
Db 2883 TAGGTGCTCTTAGTCTTTTGTATTAATAAACCTTTGGCAAACTCGAATGTTGTTAG 2942
QY 205 ----- 205
Db 2943 CAACCAAGCTATTTTCTCTCTTTGACCAACCGCTATTCTACTTGTTTGGTCGGT 3002
QY 205 ----- 205
Db 3003 GAGGCCGGGTATTTTTGTGATCCGTGAGCCCATTTTGTGCTGGGTCTCACATCC 3062
QY 205 ----- 205
Db 3063 AAGGTTCCTGGGTTCCGTTATGATTCAGACTGACGCTGAACATCTGCGCTCGCCCTGCT 3122
QY 205 ----- 205
Db 3123 TCTTGTGTGCTCTGCTATTCTGCTATTGTCATTGTAAGCGGATCCAAACACAGAACA 3182
QY 205 ----- 205
Db 3183 AAGACCCAGTCCAGATCCATCCAGATTATGTGTTTCGGCTTGCTCTGCGAGTCCACT 3242
QY 205 ----- 205
Db 3243 TTGGGTCACTGTTGCTGGCAGCAGCTCCGTGAACGTCAGCTGTGGGCTCTGTCAGGAG 3302
QY 205 ----- 205
Db 3303 GCTCAGGCTGCTCCAGCGGGACACAGTTTGGGCTCTGAAACACAGCTGAAGAAGGC 3362
QY 206 -----GlyCyspro 208
Db 3363 AGAGTGTCTTGTGATCTTTTAAATCAATGTTTTCACATTTTTCATCTCTAGGCTGCCG 3422
QY 209 ProHisTrpLysAsnPhetrAspLysCysTyrTyrPheSerValGluLysGluIlePhe 228

Db 3423 CCTCACTGGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTT 3482
QY 229 GluAspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArg 248
Db 3483 GAGGATGCAAAAGCTTTCTGTGAAGACAAGTCTTCACATCTTGTGTTTTCATAAACACTAGA 3542
QY 249 GluGluGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeu 268
Db 3543 GAGGAACAGCAATGATATAAAAAACAGATGGTAGGAGAGAGAGCCACTGGATCGGCCTC 3602
QY 269 ThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLys 288
Db 3603 ACAGACTCAGACGCGTAAATGAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAA 3662
QY 289 AsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCys 308
Db 3663 AATTGGAAGCTGGACAGCCGATTAACGGGTCTATGGCCATGGCCAGAGAAGACTGT 3722
QY 309 AlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAspPhe 328
Db 3723 GCTGGTGTGATTATGCTGGCAGTGGAAACGATTTCCAATGTGAAGACGCTCAATAACTTC 3782
QY 329 IleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 342
Db 3783 ATTTGCAAAAAGACAGGGAGACAGTACTGTCTATGTCATTA 3824
RESULT 10
AC112416 130763 bp DNA linear HTG 13-JUL-2002
LOCUS Rattus norvegicus clone CH230-350K4, *** SEQUENCING IN PROGRESS
DEFINITION *** 28 unordered pieces.
AC112416
VERSION AC112416.3 GI:21743403
KEYWORDS HTG; HTGS-PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 130763)
AUTHORS Murzyn D.M., Adams C., Adio-Oduola B., Ali-osman, F.R., Allen C.,
Albrooks S.L., Amaratunge H.C., Are J.R., Ayale, M., Banks T.,
Barbakra J., Benton J., Bimaga K., Blankenburg K., Bonnin D.,
Bouck J., Bowie S., Briefe M., Brown E., Brown M., Bryant N.P.,
Buhay C., Burck P., Burkett C., Burrell K.L., Byrd N.C.,
Carron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,
Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,
Davila M.L., Davis C., Davy Carroll L., Dederich D.A.,
Delaney K.R., Delgado O., Denn A.L., Ding Y., Din H.H.,
Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J.,
Earnhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,
Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,
Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,
Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J.,
Hernandez O., Hodgson A., Hogues M., Holloway C., Hollins B.,
Honsi F., Howard S., Huber J., Hulyk S., Hume J., Jackson L.E.,
Jacobson B., Jia Y., Johnson R., Jollivet S., Joudah S.,
Karlssoen E., Kelly S., Khan U., King L., Korvah J., Kovar C.,
Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C., Lewis L.,
Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W., Louised H.,
Lozados R., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
Mashwari M., Mapua P., Martin R., Martindale A., Martinez E.,
Massey E., Mawhney E., McLeod M.P., Meador M., Mei G., Metker M.,
Miner G., Miner Z., Mitchell T., Mohabbat K., Morgan M., Morris S.,
Moser N., Neal D., Newton J., Newton N., Nguyen M., Nguyen N.,
Nguyen N., Nickerson E., Nwokenkwo S., Ogih M., Okwuonu G.,
Oragunye N., Oviedo R., Pace A., Payton B., Peery J., Perez L.,
Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y.,
Rives M., Rojas A., Rojibokan I., Rolfe M., Ruiz S., Savary G.,
Scherer S., Scott G., Shen H., Shoostari N., Sisson I.,

Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleciyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 130763)
Worley, K.C.

Direct Submission
Submitted (21-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 130763)
Worley, K.C.

Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:20303233.

* 23699 26125: contig of 2427 bp in length
* 26126 26225: gap of unknown length
* 26226 26263: contig of 4038 bp in length
* 30264 30363: gap of unknown length
* 30364 32964: contig of 2601 bp in length
* 32965 33064: gap of unknown length
* 33065 36455: contig of 3391 bp in length
* 36456 36555: gap of unknown length
* 36556 40431: contig of 3876 bp in length
* 40432 40531: gap of unknown length
* 40532 42635: contig of 2104 bp in length
* 42636 42735: gap of unknown length
* 42736 46230: contig of 3495 bp in length
* 46231 46331: gap of unknown length
* 46331 49593: contig of 3263 bp in length
* 49594 49693: gap of unknown length
* 49694 55036: contig of 5343 bp in length
* 55037 55136: gap of unknown length
* 55137 60936: contig of 5800 bp in length
* 60937 61036: gap of unknown length
* 61037 69030: contig of 7994 bp in length
* 69031 69130: gap of unknown length
* 69131 79184: contig of 10054 bp in length
* 79185 79284: gap of unknown length
* 79285 91807: contig of 12523 bp in length
* 91808 91907: gap of unknown length
* 91908 106422: contig of 14515 bp in length
* 106423 106522: gap of unknown length
* 106523 119310: contig of 12788 bp in length
* 119311 119410: gap of unknown length
* 119411 130763: contig of 11353 bp in length.

FEATURES
Location/Qualifiers
1. 130763
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-350K4"

BASE COUNT 38110 a 25956 c 25888 g 38088 t 2721 others
ORIGIN

Alignment Scores:
Pred. No.: 3,81e-26 Length: 130763
Score: 1021.50 Matches: 215
Percent Similarity: 41.47% Conservative: 16
Best Local Similarity: 38.60% Mismatches: 37
Query Match: 52.65% Indels: 289
DB: 2 Gaps: 5

US-09-763-712A-2_COPY_206_547 (1-342) x AC112416 (1-130763)

QY 38 PheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGln 57
Db 13787 TTTCYTTATGTTTAGTCTCTGCCCCACAGGTCACAGAGGATCTCAG 13846

QY 58 GlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyPro 77
Db 13847 GGACCGCTGTGTCACAAAGGACAGAGGACAGAGGAGGAGGAGCTGTGCTCA 13906

QY 78 ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGly 97
Db 13907 CCTGGCCCTGCGGGTGAGCGGACCAATGGACAGTAGGCCCTCTCGAGAGCGTGC 13966

QY 98 GlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysPro 117
Db 13967 AGCAAGGATCAAGGCTCACAGGTCACAAAGGATCATCGTGGTCCCGAGAGACCT 14026

QY 118 GlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeu 137
Db 14027 GSCCCTCAAGGGCCCTAGTGGAGACCTGACCCACAGGTCACAGGAGGAGGAGCT 14086

QY 138 ProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGly 157
Db 14087 CCTGGCCCTCAGGGCCCTCTGCTTCCAGGACTACAGGGGCACTGTGGGTGAACCTGGA 14146

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: this is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1103: contig of 1103 bp in length
1104 1203: gap of unknown length
1204 3068: contig of 1865 bp in length
3069 3168: gap of unknown length
3169 4755: contig of 1587 bp in length
4756 4855: gap of unknown length
4856 7046: contig of 2191 bp in length
7047 7146: gap of unknown length
7147 8651: contig of 1505 bp in length
8652 8751: gap of unknown length
8752 10001: contig of 1250 bp in length
10002 10101: gap of unknown length
10102 11278: contig of 1177 bp in length
11279 11379: gap of unknown length
11379 13651: contig of 2273 bp in length
13652 13751: gap of unknown length
13752 15942: contig of 2191 bp in length
15943 16042: gap of unknown length
16043 17619: contig of 1577 bp in length
17620 17719: gap of unknown length
17720 20301: contig of 2582 bp in length
20302 20401: gap of unknown length
20402 23598: contig of 3197 bp in length
23599 23698: gap of unknown length

QY 158 valProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys 177
|||||
Db 14147 GTACCTGGACCTCGGGGCTACCAAGCTTACCAGGGTCCAGGCATGCTGGGCCAAG 14206
QY 178 GlyProGlyProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGlu 197
|||||
Db 14207 GGACCACTTGGCCCCCAGGTCCTCTCAGGAGCAATGGAGCCGTGGCTCTGCAGAAATGA 14266
QY 198 ProThrProAlaProGluAspAsn 205
|||||
Db 14267 CCAACCCAGCATCAGAGGTCAACGGTAAGTCCAAATCATCTTTTCAGCCAGAAATAAG 14326
QY 205 205
Db 14327 GCAGTGTGTAGGCTTGGCTTAACACAGACCTCAGGAGTCCCTAAATTACACTTAATATCAG 14386
QY 205 205
Db 14387 GGACCTATGTCTCTCTGGATGCTCGGTGCTGTCACTCAGGACCCGTGTCTCTGCT 14446
QY 205 205
Db 14447 TAAACATTATCAGGACTTACTCTATTAGATGAGCAGGCCCTCAGAAGATACCTATACC 14506
QY 205 205
Db 14507 CCAAGTACTATTTCAGAGCTATCTGCACCGGCTGCTCCCTGAACCATCTCTCTGTCTT 14566
QY 205 205
Db 14567 TCTTCTTTCTCTAGAACAGACCATTCCATCAATTAACCTCATGATGAGATGGGG 14626
QY 205 205
Db 14627 TTTCTGTGATAGACCACACTCTGGTTTGTGTTAAGAAGCTGCTGTGCTCTCTTTT 14686
QY 205 205
Db 14687 GACCATTAAAGTCTTCACAAACTCTGTCTAGTTAGGAACAAGCCATGCGGTCACTCA 14746
QY 205 205
Db 14747 TGCTGTGTCATGNACTTTGGTCTGAGTACACTTGTTTACAGCTCCAGCTGATCCAAG 14806
QY 205 205
Db 14807 ACTGACAATGAGTATCAGGGCCACCCACCTTCTTCTCTCTCATATAAAGCCTAAC 14866
QY 205 205
Db 14867 CCAACCAATCAGACGCTTTCCACACTGTGTGGTGTCTACTATTGTACTATAAATCC 14926
QY 205 205
Db 14927 ACNACTGTGAACATCTGTCCTATGTTCTGTACAGGGACACTCCATGGCATGCAAGCAGTG 14986
QY 205 205
Db 14987 TGAGGCTGTCACCTTGTGTATCTTAACATAGCTGAGGAAATGCAGAGTCCCTTTTTTTT 15046
QY 206 205
Db 15047 AGATCAACATTTAAACAGCTTACATTCCTPAGGATGTCGCCCTCACTGGAGAACTTCACA 15106
QY 216 AspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCys 235
|||||
Db 15107 GATAATGTACTATTTTTCGGTGGGAAAGAAATTTTGAAGATGCTAAGCTTTTCTGT 15166
QY 236 GluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGln 251
|||||
Db 15167 GAAGACAAATCTTCCCATCTCGTTTTCATATAAACTCAAGAGAGGAACAGGTATGTACGCTG 15226
QY 252 ---GlnTrpIleLysLysGlnMetValGlyArg-----GluserHisTrpIleGly 267

Db 15227 CATACTTGGTTGGAGAAAACATAATTTAGAGATAAGTTATATCTTGAATAATTCATGCGCT 15286
QY 268 LeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyr 287
|||||
Db 15287 TTGAGGAGTCGAAGAGATCAAAAATTCATGATCTCTATAACAAGGAAGCTCAGTTTG 15346
QY 288 LysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAsp 307
|||||
Db 15347 -----CATGGCTTTTATCCAGCATCGGA 15370
QY 308 CysAlaGlyLeu-----IleTyrAlaGlyGlnTrp 317
|||||
Db 15371 ATATGTGAATGTGTCATGCTTCTTAGGCCACAGATATAGTCCACCTGG 15421
RESULT 11
AC114677
LOCUS AC114677
DEFINITION Mus musculus clone RP24-213K19, WORKING DRAFT SEQUENCE, 10 ordered
AC114677
VERSION AC114677.3 GI:21362159
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 182029)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP24-213K19
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 182029)
REFERENCE 1 (bases 1 to 182029)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,
Landers,T., Lehoczký,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,C.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 182029)
REFERENCE 1 (bases 1 to 182029)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehoczký,J., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
Meneus,C., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,


```
Db 46887 TGTTCTGCTTAGACATATTACAGTCTCTACTCACTTTAGATGAGCCAGCGCTTCAGAAGA 46946
QY 203 ----- 203
Db 46947 CCCTGTACCCCAATACCTCTTCAAGATATTCTACACTGGCTGCTCCCTGACCATCCT 47006
QY 203 ----- 203
Db 47007 CCTTCATCTTTCTCTTTCTCTCTCTAGAAACAGACCATTATATCAGTTAACTCCACCAGTA 47066
QY 203 ----- 203
Db 47067 TGCAGATAGGGTTTCTGTGACAGAGTCCACATCTCTGGTTTTGCTTAAGAAGCTGTGTGG 47126
QY 203 ----- 203
Db 47127 TGCCCTTCTGTCTCTTTGACAGAAAGCTTTTGCAAAACTATCTAGTAGTAGGAGCCAAAGC 47186
QY 203 ----- 203
Db 47187 GTGTGGTCCCTCATGCTGGTCCACAAGCCAGACTTTTGGTTCTGGAAGTGTGCTTAFTTAC 47246
QY 203 ----- 203
Db 47247 AACCTCCAGCGGATCCAGCCGACAGTGTATCAGAACCCACCACCTTTCTTTTCTGT 47306
QY 203 ----- 203
Db 47307 CGTCATAAGCCTAGCCCAACCCAGTCCACCTTTCTTCTGCTGTGGTGTCTACATGCTG 47366
QY 203 ----- 203
Db 47367 TGTGCTGTACAGTCCACATCTGAGAACATCTGCCGGGTTCTACACAGGACACCCATTG 47426
QY 203 ----- 203
Db 47427 GCATGCAAGTGTACAAGCTGTGACTTTGATATCTTAACACACAGCTGAGGAATGCAGG 47486
QY 204 ----- AspAsnGlyCysProPro 209
Db 47487 GTTCCTTCTGTGTTTTTAATCAACAGTAAACTTTTACATTCCTA-CGATGTCGGCT 47545
QY 210 HisTrpLysAsnPhetThrAspLysCysTyrTrpSerValGluLysGluLeuGlu 229
Db 47546 CACTGGAAGAACTTCACAGATAAATGCTACTATTTTTCATTGGAAGAAATTTTGAA 47605
QY 230 AspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArgGlu 249
Db 47606 GATGCTAAGCTTTTCTGTGAAGACAAATCTCCCATCTCGTTTTCATAAAGCTCAAGAGAA 47665
QY 250 GluGlnGln-----TrpIleLysLysGlnMetValGlyArgGlu 262
Db 47666 GAACAGGT-ATGATATGTGAGCTGCACACTTGGTGGAGAAACTATATTATAGATA- 47723
QY 263 SerHisTrpIleClyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGly 282
Db 47724 AGTTATACTTGGAAATGTGATGCTCTTTGAGAGAGTCAAGAGATCAA----- 47771
QY 283 ThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHis 302
Db 47772 -----AATTCAATGGTCTCATAAACAA 47792
QY 303 GlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrp---AsnAspPheGln 321
Db 47793 AGGAAGCGTGCATTTGCGATGGCTTTTCCAGGATGAGAGTATGATGATGAATCAAA 47852
RESULT 12
AC102618
LOCUS AC102618 193208 bp DNA linear HTG 21-AUG-2002
DEFINITION Mus musculus clone RP23-426G16, WORKING DRAFT SEQUENCE, 11
unordered pieces.
AC102618
```

VERSION
KEYWORDS
SOURCE
ORGANISM

AC102618.2 GI:22381604

HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

house mouse.

Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 193208)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-426G16

REFERENCE
AUTHORS

2 (bases 1 to 193208)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campotiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cook,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kanat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., MCPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye.W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

TITLE
JOURNAL

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 193208)

REFERENCE
AUTHORS

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

TITLE
JOURNAL

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17061704.

COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L19116

Center clone name: 426_G16

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Db 98900 GTGTGGTCCCTCATGCTGGTCCCAAGCCAGCACTTTGGTCTGAAGTGTGCTATTATAC 98959
QY 203 ----- 203
Db 98960 AACCTCCAGCCGATCCAGCCGACAGTGTATCAGAACCCACCCACTTCTTTTTCGTGT 99019
QY 203 ----- 203
Db 99020 CGTCATAAGCCCTAGCCCAACCCAGTCCACACCTTCTTCACTGTGTGGTGTCACTGCTG 99079
QY 203 ----- 203
Db 99080 TGTGCTGTACATCCACATCTGAGAACATCTGCCGGGTTCTACACAGGACACCCATTG 99139
QY 203 ----- 203
Db 99140 GCATCAAGTGTACAGGCTGTGACTTGTATATCTTAACACAGCTGAGGAATGCGAGG 99199
QY 204 ----- AspAsnGlyCysProPro 209
Db 99200 GTTCTCTTGTGTTTTTAAATCAACAGTTAAACTTTTACATTCCTA-GGATGTCGCGCT 99258
QY 210 HistTrpLysAsnPhetAspLysCysTyrrPheSerValGluLysGluIlePheGlu 229
Db 99259 CACTGGAGAACTTCACAGATAAATGCTACTATTTTTCATTGCAAAAAGAAATTTTGAA 99318
QY 230 AspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGlu 249
Db 99319 GATGCTAAGCTTTCTGTGAACACAAATCTCCCATCTCGTTTTCATAAATCAACACAGAA 99378
QY 250 GluGlnGln-----TrpIleLysLysGlnMetValGlyArgGlu 262
Db 99379 GAACAGGT-ATGATATGTGAGTGCACACTTGTGGGAAACATATAATTTAGAGATA- 99436
QY 263 SerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGly 282
Db 99437 AGTTATACTTGGAAATGATGCCTTTTGAGAGAGTCAAGATGATCAAA----- 99484
QY 283 ThrSerProAspTyrrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHis 302
Db 99485 -----AATCATGGTCTCTATAACAA 99505
QY 303 GlyProGlyGluAspCysAlaGlyLeuIleTyrrAlaGlyGlnTrp---AsnAspPheGln 321
Db 99506 AGGAAGGCTGCAGTTTGCATGGGCTTTTCCAGGTATGAGAGTATGATGAATGATCAAA 99565
RESULT 13
AC016128/c
LOCUS Homo sapiens chromosome 18 clone RP11-324G2 map 18, WORKING DRAFT
DEFINITION Homo sapiens chromosome 18 clone RP11-324G2 map 18, WORKING DRAFT
SEQUENCE, 8 unordered pieces.
AC016128
AC016128.4 GI:10046526
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169088)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boquslavkiy,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collumore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Doneian,L., Doyle,M.,
Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,

Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 9, 2000 this sequence version replaced gi:5649269.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 1999
Center clone name: 324_G2
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 16522 bases at least Q40
Consensus quality: 167248 bases at least Q30
Consensus quality: 167996 bases at least Q20
Insert size: 177000; agarose-fp
Quality coverage: 6.2 in Q20 bases; agarose-fp
Quality coverage: 6.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 17943: contig of 17943 bp in length
* 17944 18043: gap of 100 bp
* 18044 22424: contig of 4381 bp in length
* 22425 22524: gap of 100 bp
* 22525 38094: contig of 15570 bp in length
* 38095 38194: gap of 100 bp
* 38195 49220: contig of 11026 bp in length
* 49221 49320: gap of 100 bp
* 49321 66394: contig of 17074 bp in length
* 66395 66494: gap of 100 bp
* 66495 91692: contig of 25198 bp in length
* 91693 91792: gap of 100 bp
* 91793 120869: contig of 29077 bp in length
* 120870 120969: gap of 100 bp
* 120970 169088: contig of 48119 bp in length.
FEATURES
source
1..169088
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-324G2"
/clone_lib="RP11-11 Human Male BAC"
1..17943
/note="assembly_fragment"
clone_end:SP6
vector_side:left
18044..22424
/note="assembly_fragment"
22525..38094
/note="assembly_fragment"
38195..49220
/note="assembly_fragment"


```
misc_feature 49321..66394
/note="assembly_fragment"
misc_feature 66495..91692
/note="assembly_fragment"
misc_feature 91793..120869
/note="assembly_fragment"
misc_feature 120970..169088
/note="assembly_fragment"
clone_end:T7
vector_size:right"
BASE COUNT 48103 a 33537 c 34088 g 52660 t 700 others
ORIGIN
Alignment Scores:
Pred. No.: 8.52e-24 Length: 169088
Score: 954.50 Matches: 191
Percent Similarity: 58.40% Conservative: 21
Best Local Similarity: 52.62% Mismatches: 53
Query Match: 49.20% Indels: 98
DB: 2 Gaps: 8
US-09-763-712A-2_COPY_206_547 (1-342) x AC016128 (1-169088)
Qy 16 AsnLeuSerValIleMetGluMetLysLeuValAspSerLysHisGly-----Gln 33
||||| : : : : : |||||
Db 153903 AATTGGAGATTATTATAAGAGAAAAATTATTCATAACTATCATTCACAGTGGATTTTCAT 153844
Qy 34 LeuIleLysAsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAsp 53
: : : : : |||||
Db 153843 GTTGACACTTTTCTTTTATGCTTTAGTCCACCGGCCCCAGGGGTCGAAGAGTGAC 153784
Qy 54 ArgGlySerGlnGlyProGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGly 73
||||| : : : : : |||||
Db 153783 AGAGGATCCAGGGACCCCTGGCCCAACTGCAACAAGGACAGAAAGAGAGAGGGG 153724
Qy 74 GluProGlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProPro 93
||||| : : : : : |||||
Db 153723 GAGCGTGACACCACTGGCCCTCGGGGTGAGAGAGCCCAATTGGACCACTGGTCCCCC 153664
Qy 94 GlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySer 113
||||| : : : : : |||||
Db 153663 GGAGAGCGTGGCGGCAAGGATCTAAGGCTCCCAAGGCCCAAGGCTCCCGTGGTTCC 153604
Qy 114 ProGlyLysProGlyProGlnGlyProSerGlyAspProGlyProGlyProProGly 133
||||| : : : : : |||||
Db 153603 CCTGGGAAGCCCGGCCCTCAGGCTCCAGTGGGGACCCAGGCCCCCGGCCCCACGAGC 153544
Qy 134 LysGluGlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrVal 153
||||| : : : : : |||||
Db 153543 AAAGAGGAGCTCCCGGCCCTCAGGCGCCCTCTGGCTTCCAGGGACTTCAGGGCACCGT 153484
Qy 154 GlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMet 173
||||| : : : : : |||||
Db 153483 GGGAGCGCTGGGTGCTTGGACCTCGGGACTCGCAGGCTTGGCGGGGTACCGAGCATG 153424
Qy 174 ProGlyProLysGlyProGlyProGlyProGlyProSerGlyAlaValValProLeuAla 193
||||| : : : : : |||||
Db 153423 CCAGSCCCCAAGGGCCCCCGGCCCTCTTGCCCATCAGGAGCGGTGGTGGCCCTGGCC 153364
Qy 194 LeuGlnAsnGluProThrProAlaProGluAspAsnGly----- 206
||||| : : : : : |||||
Db 153363 CTGCAAGATGAGCCCAACCCAGCACCGGAGGACAAATGTAAGTCAAGCCCTCCCTCCAC 153304
Qy 207 -----CysPro-ProH 210
||||| : : : : : |||||
Db 153303 CCAGGGGACAGGGCAGTGCATGCTTGACTGAGCAACCCCACTGGTGGCCCA 153244
Qy 210 sTriPlysAsnPheThrAspLysCysTyTyTyPheSerValGluLysGluIlePheGluAs 230
||||| : : : : : |||||
Db 153243 C-----CCATGTTATTT----- 153231
Qy 230 pAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGlu 250
||||| : : : : : |||||
```


* 2385 3056: contig of 672 bp in length
* 3057 3156: gap of 100 bp
* 3157 3851: contig of 695 bp in length
* 3852 3951: gap of 100 bp
* 3952 4627: contig of 676 bp in length
* 4628 4727: gap of 100 bp
* 4728 5409: contig of 682 bp in length
* 5410 5509: gap of 100 bp
* 5510 6183: contig of 674 bp in length
* 6184 6283: gap of 100 bp
* 6284 6984: contig of 701 bp in length
* 6985 7084: gap of 100 bp
* 7085 7768: contig of 684 bp in length
* 7769 7868: gap of 100 bp
* 7869 8550: contig of 682 bp in length
* 8551 8650: gap of 100 bp
* 8651 9352: contig of 702 bp in length
* 9353 9452: gap of 100 bp
* 9453 10140: contig of 688 bp in length
* 10141 10240: gap of 100 bp
* 10241 10922: contig of 682 bp in length
* 10923 11022: gap of 100 bp
* 11023 11709: contig of 687 bp in length
* 11710 11809: gap of 100 bp
* 11810 12511: contig of 702 bp in length
* 12512 12611: gap of 100 bp
* 12612 13304: contig of 693 bp in length
* 13305 13404: gap of 100 bp
* 13405 14095: contig of 691 bp in length
* 14096 14195: gap of 100 bp
* 14196 14881: contig of 686 bp in length
* 14882 14981: gap of 100 bp
* 14982 15674: contig of 693 bp in length
* 15675 15774: gap of 100 bp
* 15775 16460: contig of 686 bp in length
* 16461 16560: gap of 100 bp
* 16561 17250: contig of 690 bp in length
* 17251 17350: gap of 100 bp
* 17351 18040: contig of 690 bp in length
* 18041 18140: gap of 100 bp
* 18141 18836: contig of 696 bp in length
* 18837 18936: gap of 100 bp
* 18937 19621: contig of 685 bp in length
* 19622 19721: gap of 100 bp
* 19722 20425: contig of 704 bp in length
* 20426 20525: gap of 100 bp
* 20526 21224: contig of 699 bp in length
* 21225 21324: gap of 100 bp
* 21325 22017: contig of 693 bp in length
* 22018 22117: gap of 100 bp
* 22118 22815: contig of 698 bp in length
* 22816 22915: gap of 100 bp
* 22916 23609: contig of 694 bp in length
* 23610 23709: gap of 100 bp
* 23710 24402: contig of 693 bp in length
* 24403 24502: gap of 100 bp
* 24503 25191: contig of 689 bp in length
* 25192 25291: gap of 100 bp
* 25292 25981: contig of 690 bp in length
* 25982 26081: gap of 100 bp
* 26082 26774: contig of 693 bp in length
* 26775 26874: gap of 100 bp
* 26875 27563: contig of 689 bp in length
* 27564 27663: gap of 100 bp
* 27664 28355: contig of 692 bp in length
* 28356 28455: gap of 100 bp
* 28456 29145: contig of 690 bp in length
* 29146 29245: gap of 100 bp
* 29246 29943: contig of 698 bp in length
* 29944 30043: gap of 100 bp
* 30044 30745: contig of 702 bp in length
* 30746 30845: gap of 100 bp
* 30846 31535: contig of 690 bp in length

* 31536 31635: gap of 100 bp
* 31636 32329: contig of 694 bp in length
* 32330 32429: gap of 100 bp
* 32430 33114: contig of 685 bp in length
* 33115 33214: gap of 100 bp
* 33215 33898: contig of 684 bp in length
* 33899 33998: gap of 100 bp
* 33999 34694: contig of 696 bp in length
* 34695 34794: gap of 100 bp
* 34795 35486: contig of 692 bp in length
* 35487 35586: gap of 100 bp
* 35587 36292: contig of 706 bp in length
* 36293 36392: gap of 100 bp
* 36393 37095: contig of 703 bp in length
* 37096 37195: gap of 100 bp
* 37196 37872: contig of 677 bp in length
* 37873 37972: gap of 100 bp
* 37973 38663: contig of 691 bp in length
* 38664 38763: gap of 100 bp
* 38764 39449: contig of 686 bp in length
* 39450 39549: gap of 100 bp
* 39550 40236: contig of 687 bp in length
* 40237 40336: gap of 100 bp
* 40337 41025: contig of 689 bp in length
* 41026 41125: gap of 100 bp
* 41126 41818: contig of 693 bp in length
* 41819 41918: gap of 100 bp
* 41919 42608: contig of 690 bp in length
* 42609 42708: gap of 100 bp
* 42709 43399: contig of 691 bp in length
* 43400 43499: gap of 100 bp
* 43500 44169: contig of 670 bp in length
* 44170 44269: gap of 100 bp
* 44270 44966: contig of 697 bp in length
* 44967 45066: gap of 100 bp
* 45067 45750: contig of 684 bp in length
* 45751 45850: gap of 100 bp
* 45851 46545: contig of 695 bp in length
* 46546 46645: gap of 100 bp
* 46646 47336: contig of 691 bp in length
* 47337 47436: gap of 100 bp
* 47437 48123: contig of 687 bp in length
* 48124 48223: gap of 100 bp
* 48224 48899: contig of 676 bp in length
* 48900 48999: gap of 100 bp
* 49000 49672: contig of 673 bp in length
* 49673 49772: gap of 100 bp
* 49773 50465: contig of 693 bp in length
* 50466 50565: gap of 100 bp
* 50566 51254: contig of 689 bp in length
* 51255 51354: gap of 100 bp
* 51355 52062: contig of 708 bp in length
* 52063 52162: gap of 100 bp
* 52163 52847: contig of 685 bp in length
* 52848 52947: gap of 100 bp
* 52948 53646: contig of 699 bp in length
* 53647 53746: gap of 100 bp
* 53747 54409: contig of 663 bp in length
* 54410 54509: gap of 100 bp

Alignment Scores:

Pred. No.: 2.09e-23 Length: 71044
Score: 935.50 Matches: 165
Percent Similarity: 92.43% Conservative: 6
Best Local Similarity: 89.19% Mismatches: 11
Query Match: 48.22% Indels: 3
DB: 2 Gaps: 1

US-09-763-712a-2_COPY_206_547 (1-342) x AC024368 (1-71044)

Qy 25 LysLeuValAspSerLysHisGlyGlnLeuLeu-----LysAsnPhetrIleLeu 41
|||||: : : ||| ||| : : :
Db 43544 AAATATTATCACTAATCAATTCACAGTGGATTTCATGTTGACGGGGGGCTTTTAGCA 43603

QY 42 GlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGly 61
Db 43604 AAAAGTCCACCGGGCCCGAGGGTCCACAGAGGTGACAGAGGATCCACAGGACCCCTGGC 43663
QY 62 ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAla 81
Db 43664 CCAACTGGCAACAAGGACAGAAAGAGAGAGAGGGGAGCCTGGACCACTGGCCCTGGC 43723
QY 82 GlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySer 101
Db 43724 GGTGAGAGAGCCCGAGGGGACCACTGTCTCCCGGAGAGCGTGGCGGCAAGGATCT 43783
QY 102 LysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGly 121
Db 43784 AAAGGCTCCAGGGCCCAAGAGCTCCCGTGGTTCCTCTGGGAAGCCCGCCCTCAGGCGC 43843
QY 122 ProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGln 141
Db 43844 TCCAGTGGGGACCCAGGGCCCGGGCCCAAGAGGAGGACTCCCGGGCCCTCAG 43903
QY 142 GlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyPro 161
Db 43904 GGCCCTCCTGGCTTCAGGGACTTCAGGGCACCGTTGGGGAGCCTGGGTGGCTGGACCT 43963
QY 162 ArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGly 181
Db 43964 CGGGGACTGCCAGGCTTGCTGGGTACCAAGGCATGCCAGGCCCAAGGGCCCGCCCGGC 44023
QY 182 ProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAla 201
Db 44024 CCTCTGGCCCAATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCCAAACCCAGCA 44083
QY 202 ProGluAspAsnGly 206
Db 44084 CCGAGGACAATGGT 44098

Search completed: March 21, 2003, 11:21:28
Job time : 2501.11 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

SUMMARIES

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 02:20:06 ; Search time 179,385 Seconds
(without alignments)
4293.470 Million cell updates/sec

Title: US-09-763-712a-2_COPY_206_547
Perfect score: 1940
Sequence: 1 MQQDLMSRLDTEVANLSVI.....EDVNNFICKDRETVLSSAL 342

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n_model -DEV=xlp
-Q=cqn2_l1/USPRO-spool/US09763712/runat_14032003_100949_18091/app_query.fasta_1.1877
-DB=N_Geneseq_101002 -OFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712.ecgn_l_1.626 @runat_14032003_100949_18091 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRGADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|------------------------------|
| 1 | 1940 | 100.0 | 2024 | 21 | AAA07697 Human collectin en |
| 2 | 1940 | 100.0 | 2262 | 24 | ABA97932 Human scavenger re |
| 3 | 1940 | 100.0 | 2318 | 22 | AAI58842 Human polynucleoti |
| 4 | 1940 | 100.0 | 2628 | 22 | AAH43036 Nucleotide sequenc |
| 5 | 1934 | 99.7 | 2005 | 24 | ABL95574 Human anglogenesis |
| 6 | 1934 | 99.7 | 2005 | 24 | ABL88085 Human polynucleoti |
| 7 | 1934 | 99.7 | 2181 | 22 | AAI60628 Human PRO7223 CDNA |
| 8 | 1934 | 99.7 | 2641 | 22 | AAI60628 Human EXWAD-14 cod |
| 9 | 1934 | 99.7 | 2929 | 24 | ABQ92072 Human polynucleoti |
| 10 | 1934 | 99.7 | 2930 | 20 | AAV55746 Human secreted pro |
| 11 | 1801 | 92.8 | 2637 | 22 | AAH43037 Nucleotide sequenc |
| 12 | 1715.5 | 88.4 | 1521 | 23 | AAH71133 DNA encoding novel |
| 13 | 1158 | 59.7 | 2256 | 22 | AAH43054 Nucleotide sequenc |
| 14 | 540 | 27.8 | 3685 | 20 | AAH27858 Human CSR3 protein |
| 15 | 540 | 27.8 | 3810 | 20 | AAH27856 Human CSR1 protein |
| 16 | 474.5 | 24.5 | 1410 | 24 | ABK84622 Human CDNA differe |
| 17 | 474.5 | 24.5 | 1410 | 24 | ABL69664 Lung cancer relate |
| 18 | 471 | 24.3 | 5676 | 15 | AAQ64556 Human collagen (Ty |
| 19 | 471 | 24.3 | 8284 | 22 | AAAL26526 Human breast cance |
| 20 | 471 | 24.3 | 8284 | 22 | AAAL26553 Human breast cance |
| 21 | 471 | 24.3 | 8284 | 22 | AAAL26600 Human breast cance |
| 22 | 462 | 23.8 | 4428 | 22 | AAAD06574 Bovine alphas(III) |
| 23 | 462 | 23.8 | 4428 | 22 | AAAD06575 Bovine alphas(III) |
| 24 | 459.5 | 23.7 | 756 | 14 | AAQ43034 Collagen-like poly |
| 25 | 459.5 | 23.7 | 756 | 17 | AAAT16768 Collagen-like poly |
| 26 | 458.5 | 23.6 | 4821 | 23 | AAAS86866 DNA encoding novel |
| 27 | 454 | 23.4 | 3171 | 21 | AAA12503 cDNA encoding a hu |
| 28 | 454 | 23.4 | 3174 | 21 | AAZ99843 DNA encoding human |
| 29 | 453 | 23.4 | 6158 | 24 | ABL62095 Colon adenocarcino |
| 30 | 453 | 23.4 | 6158 | 24 | ABL63452 Lung cancer relate |
| 31 | 453 | 23.4 | 6158 | 24 | ABK35486 Human endometrial |
| 32 | 452.5 | 23.3 | 3170 | 21 | AAA12493 cDNA encoding huma |
| 33 | 452.5 | 23.3 | 3171 | 21 | AAA12502 cDNA encoding a hu |
| 34 | 452.5 | 23.3 | 3181 | 19 | AAV59358 Nucleotide sequenc |
| 35 | 452.5 | 23.3 | 3349 | 17 | AAAT16518 Collagen Al/decori |
| 36 | 452.5 | 23.3 | 3349 | 21 | AAA12500 cDNA encoding a ch |
| 37 | 452.5 | 23.3 | 3531 | 21 | AAA12497 cDNA encoding a ch |
| 38 | 452.5 | 23.3 | 3535 | 17 | AAAT16515 Collagen Al/tgf-be |
| 39 | 452.5 | 23.3 | 3541 | 17 | AAAT16516 cDNA encoding a ch |
| 40 | 452.5 | 23.3 | 3541 | 21 | AAAL2498 Collagen Al/tgf-be |
| 41 | 452.5 | 23.3 | 4192 | 17 | AAAT16517 Collagen Al/decori |
| 42 | 452.5 | 23.3 | 4409 | 19 | AAV60814 Human recombinant |
| 43 | 452.5 | 23.3 | 4770 | 22 | AAAS22441 Human CDNA encodin |
| 44 | 452.5 | 23.3 | 6728 | 22 | AAF90491 Human pro-alpha-1 |
| 45 | 452.5 | 23.3 | 6728 | 24 | ABN97451 Gene #3949 used to |

ALIGNMENTS

RESULT 1
AAA07697
ID AAA07697 standard; DNA; 2024 BP.
XX AAA07697;
AC AAA07697;
XX 20-JUN-2000 (first entry)
DT 20-JUN-2000 (first entry)
XX Human collectin encoding DNA.
DE Human collectin encoding DNA.
XX Collectin; human; antibacterial; antiviral; ds.
XX Homo sapiens.
OS Homo sapiens.
XX Key
FH CDS
FT Location/Qualifiers
FT 55..1698
FT /*tag= a
FT /product= "collectin"

AC ABA97932;
XX
DT 25-APR-2002 (first entry)
XX
DE Human scavenger receptor-like protein encoding cDNA SEQ ID NO 1.
XX
KW Human; scavenger; receptor; diagnosis; treatment; autoimmune disease;
KW rheumatoid arthritis; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 33..2262
FT /*tag= a
FT /product= "scavenger receptor-like protein"
XX
PN JP2001340089-A.
XX
PD 11-DEC-2001.
XX
PF 08-DEC-2000; 2000JP-0375066.
XX
PR 27-MAR-2000; 2000JP-0090772.
XX
PA (SHIO) SHIONOGI & CO LTD.
XX
DR WPI; 2002-144965/19.
DR P-PSDB; ABB08642.
XX
PT New scavenger receptor-like protein for diagnosis, prevention and
PT treatment of autoimmune disease, such as rheumatoid arthritis -
XX
PS Claim 3; Fig 1; 38pp; Japanese.
XX
CC The invention relates to a human scavenger receptor-like protein. The
CC protein is useful as a target molecule for diagnosis, prevention and
CC treatment of autoimmune diseases such as rheumatoid arthritis.
XX
SQ Sequence 2262 BP; 670 A; 576 C; 581 G; 435 T; 0 other;

Alignment Scores:
Pred. No.: 1,74e-58 Length: 2262
Score: 1940.00 Matches: 342
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0

US-09-763-712A-2_COPY_206_547 (1-342) x ABA97932 (1-2262)

QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1234 ATGCAACAAGATTTCATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATT 1293
QY 21 MetGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhethrIle 40
Db 1294 ATGGAAGAATGAAGCTAGTACTCAAGCATGGTGGTCAATCAAGAAATTTACAATA 1353
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyProArgGlySerGlnGlyPro 60
Db 1354 CTACAGGTGCCACCGGGCCCCAGGGGTCCAGAGGTGACAGAGGTCCCGAGGACCCCT 1413
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyPro 80
Db 1414 GCCCAACTGGCAACAGGACAGAAAGAGAGAGAGGGGGAGCCCTGGACCCTGGCCCT 1473
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGly 100
Db 1474 CGGGGTGAGAGAGGCCCAATTGACACAGCTGCTCCCGGAGAGCGTGGCGGCAAGGA 1533
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGln 120
Db 1534 TCTAAAGGCTCCAGGGGCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAG 1593

QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlyLeuProGlyPro 140
Db 1594 GCGCCAGTGGGAGCCAGGCCCGCGGCCACAGGCAAGAGAGGACTCCCGGCCCT 1653
QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
Db 1654 CAGGGCCCTCCCTGGCTTCCAGGACTTCAGGGCAGCGTTGGGAGCCTGGGTGCTGA 1713
QY 161 ProArgGlyLeuProGlyValProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 1714 CCTCGGGGACTCCAGGCTTGCTGGGTACCAGGATGCCAGGCCCAAGGCCCCCT 1773
QY 181 GlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro 200
Db 1774 GCGCTCTCTGGCCCATCAGAGCGGTGGTGGCTGGCCCTGCAGAAATCAGCAACCCG 1833
QY 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPhethrAspLysCysTrpTyr 220
Db 1834 GCACCGGAGGACAATGGCTGCCCGCTCCTCTGGAAGAACTTCACAGACAAATGCTACTAT 1893
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
Db 1894 TTTTCAGTTGACAAAGAAATTTTTCAGGATGCAAGCTTTCTGTGAAGACAACTCTTCA 1953
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1954 CATCTGTTTTCATAAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGG 2013
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 2014 AGAGAGAGCCACTGGATCGGCTCACAGACTCAGAGCGTGAATAATGAATGGAAGTGGCTG 2073
QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 2074 GATGGACATCTCCAGACTACAAAATTGGAAGCTGGACCGCGGATTAACCTGGGTGCTAT 2133
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 2134 GGCCATGGGCCAGGAGAGACTGTGCTGGTTGATTATGCTGGGCAGTGAACGATTTC 2193
QY 321 GlnCysGluAspValAsnAsnPhelIleCysGluLysAspArgGluThrValLeuSerSer 340
Db 2194 CAATGTGAAGAGCTCAATAACTTCAATTCGTAATAAGAGAGAGGAGACAGTACTGTCTATCT 2253
QY 341 AlaLeu 342
Db 2254 GCATTA 2259

RESULT 3
AAI58842
ID AAI58842 standard; cDNA; 2318 BP.
XX
AC AAI58842;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1045.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Dermanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR P-PSDB; AAM39686.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 1045; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other;
 XX

Alignment Scores:
 Pred. No.: 1.77e-58 Length: 2318
 Score: 1940.00 Matches: 342
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0

US-09-763-712a-2_COPY_206_547 (1-342) x AAI58842 (1-2318)

QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
 DB 663 ATGCACACAGATTGTAGAGTGGAGGTAGACACTGAAGTAGCCAACTATCAGTGATT 722

QY 21 MetGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
 DB 723 ATGGAAGAAATGAAGCTAGACTCCAAAGCATGGTCAGCTCATCAAGAAATTACAAAT 782

QY 41 LeuGlnGlyProGlyProGlyProArgGlyAspArgGlySerGlnGlyProPro 60
 DB 783 CTACAAAGTCCACCGGGCCCGAGGGTCCCAAGAGGTGACAGAGGATCCCGAGGACCCCT 842

QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
 DB 843 GCCCCAACTGGCAACAGGACAGAAAGAGAGAGGGGGAGCCCTGGACCTGGCCCT 902

QY 81 AlaGlyGluArgGlyProTleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly 100
 DB 903 GCGGGTGAGAGAGGGCCCAATTGGACCACTGGTCCCGCCCGAGAGGCGTGGCGCAAGGA 962

QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
 DB 963 TCTAAAGGCTCCACAGGGCCCCAAAGGCTCCGCTGGTTCCTCCCTGGGAAGCCGCGCCCTCAG 1022

QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyPro 140
 DB 1023 GGCCCCAGTGGGACCCAGGCCCCCGGGCCACCAGGCAAGAGAGGACTCCCGGCGCCCT 1082

QY 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
 DB 1083 CAGGGCCCTCCCTGGCTTCAGGGACTTCAGGGCACCCTGGGGAGGCTGGGGTGCCTGGA 1142

QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
 DB 1143 CCTCGGGGACTGCCAGGCTTGCTGGGGTACCAGGATGCCAGGCCCAAGGCCCCCCC 1202

QY 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
 DB 1203 GGCCCTCCCTGGCCCATCAGGAGCGGTGGTGGCCCTGCAGAAATGAGCCAAACCCG 1262

QY 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyr 220
 DB 1263 GCACCGGAGGACAAATGGCTGCCCGCTCACTGGAAGAACTTCACAGCAAAATGCTACTAT 1322

QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
 DB 1323 TTTTCAGTTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTAAGCAAGTCTTCA 1382

QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
 DB 1383 CATCTTGTGTTTATATAAACAAGTACAGAGAGAACCAATGATATAAAACAGATGGTAGGG 1442

QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
 DB 1443 AGAGAGAGCCACTGGATCGGCTCACAGACTCAGACCGTGAACCAAGTGAAGTGGCTG 1502

QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
 DB 1503 GATGGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAT 1562

QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
 DB 1563 GGCCATGGCCAGGAGAGACTGTGCTGGGTGATTATTGCTGGGCGAGTGGAAACGATTTC 1622

QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
 DB 1623 CAATGTGAAGACGTCAATAACTTTCATTTCGAAAAAAGACAGGGAGACAGTACTGTCTATCT 1682

QY 341 AlaLeu 342
 DB 1683 GCATTA 1688

RESULT 4
 AAH43036
 ID AAH43036 standard; DNA; 2628 BP.
 XX
 AC AAH43036;
 XX
 DT 15-OCT-2001 (first entry)
 XX
 DE Nucleotide sequence of a human scavenger receptor.
 XX
 KW Human; scavenger receptor; SRCL-PI; macrophage; basal immunity;
 KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
 KW low density lipoprotein; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 74..2302
 FT /*tag= a
 FT /product= "scavenger receptor"
 XX


```

PN WO200159107-A1.
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-JP00874.
XX
XX 14-FEB-2000; 2000JP-0035155.
XX
XX 10-OCT-2000; 2000JP-0309069.
XX
XX (FUSO ) FUSO PHARM IND LTD.
XX
XX Wakamiya N;
XX
XX WPI; 2001-497076/54.
XX
XX P-PSDB; AAG63346.
XX
XX New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
XX useful for treatment and diagnosis of diseases associated with oxidized
XX low-density lipoprotein accumulation
XX
XX Claim 2; Page 79-84; 118pp; Japanese.
XX
XX The present sequence encodes a human scavenger receptor, designated
XX SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
XX are useful in clarifying the functions of macrophages and basal
XX immunity. They are also useful in the treatment, prevention, diagnosis
XX and investigation of diseases such as arteriosclerosis, diabetic
XX complications, bacterial infection and restenosis following angioplasty,
XX which are associated with accumulation of oxidized low density
XX lipoprotein and the binding of advanced glycation end-products into
XX cells.
XX
XX Sequence 2628 BP; 783 A; 661 C; 656 G; 528 T; 0 other;
XX
XX Alignment Scores:
XX
XX Pred. No.: 1,93e-58 Length: 2628
XX Score: 1940.00 Matches: 342
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-763-712A-2_COPY_206_547 (1-342) x AAH43036 (1-2628)
XX
XX QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
XX
XX DB 1274 ATGCAACAAGATTTCATGAGTCTGAGGTAGACACTGAAGTAGCCAACTTATCATGTATT 1333
XX
XX QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
XX
XX DB 1334 ATGGAAGAAATGAAGTAGTACACTCCCAAGCATGCTCAGCTCATCAAGAAATTTACAATA 1393
XX
XX QY 41 LeuGlnGlyProGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
XX
XX DB 1394 CTACAAAGTGTCCACCGGGGCCCAAGGGTCCAAAGAGTGCACAGAGATCCCGAGGACCCCT 1453
XX
XX QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGlyLysGlyGluProGlyProGlyPro 80
XX
XX DB 1454 GGCCCACTGGCAACAGGGACAGAAAGAGAGAGAGGGGGAGCCTGGACCCTGGCCCT 1513
XX
XX QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyProGlyGluArgGlyLysGly 100
XX
XX DB 1514 GCGGTGAGAGAGGCCCAATTTGACACAGCTGGTCCCGGGAGAGCGTGGCGGCAAGGA 1573
XX
XX QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
XX
XX DB 1574 TCTAAGGCTCCAGGGGCCCAAGGGCTCCCGTGGTCCCGTGGGAAGCCCGGCCCTCAG 1633
XX
XX QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyPro 140
XX
XX DB 1634 GGCCCACTGGGGACCCAGGCCCCCGGGCCCGCCAGGCAAGAGAGGACTCCCGGCCCT 1693
XX
XX QY 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160

```

```

Db 1694 CAGGGCCCTCCTGGCTTCCAGGGACTTCCAGGGACCGTTGGGAGCCTGGGGTCTGGA 1753
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 1754 CCTCGGGACTGCCAGGCTTGGCTGGGTACAGGCATGCCAGGCCCGCCAGGCCCGCCC 1813
QY 181 GlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro 200
Db 1814 GGCCCTCCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTCAGAAATGAGCAACCCG 1873
QY 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTrpTyr 220
Db 1874 GCACCGGAGGACAATGGCTGCCCGCTCCTCCTGGAAGAACTTCCACAGACAAATGCTACTAT 1933
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
Db 1934 TTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGACTTTCTGTGAGACAAAGCTTCA 1993
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1994 CATCTTGTTCATAAACACTAGAGAGGAACAGCAATGGATGATAAAAAACAGATGGTAGGG 2053
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 2054 AGAGAGAGCCACTGGATCGGCTCAGACTCAGAGCGTGAATAATCAATGGAAGTGGCTG 2113
QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 2114 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACCGCGATTAACCTGGGTGCTAT 2173
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 2174 GGCCATGGCCAGGAGAGACTGTCTGGTGTGATTATGCTGGCAGTGGAGCAATTTTC 2233
QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
Db 2234 CAATGTGAAGAGCTCAATAACTTTCATTTGCGAAAAAGACAGGGAGACAGTACTGTCTATCT 2293
QY 341 AlaLeu 342
Db 2294 GCATTA 2299
XX
XX RESULT 5
XX ABL95574
XX ID ABL95574 standard; cDNA; 2005 BP.
XX AC ABL95574;
XX DT 19-JUL-2002 (first entry)
XX DE Human angiogenesis related cDNA PRO7223 SEQ ID NO: 27.
XX KW Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
XX KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
XX KW cardiatic; cytostatic; antiangiogenic; hypotensive; vulnerary;
XX KW antiarteriosclerotic; gene; ss.
XX OS Homo sapiens.
XX PN WO200208284-A2.
XX PD 31-JAN-2002.
XX PF 09-JUL-2001; 2001WO-US21735.
XX PR 20-JUL-2000; 2000US-219556P.
XX PR 25-JUL-2000; 2000US-220624P.
XX PR 25-JUL-2000; 2000US-220664P.
XX PR 28-JUL-2000; 2000WO-US20710.
XX PR 02-AUG-2000; 2000US-222695P.
XX PR 17-AUG-2000; 2000US-0643657.
XX PR 23-AUG-2000; 2000WO-US23522.

```

PR 24-AUG-2000; 2000WO-US23328.
PR 17-SEP-2000; 2000US-230978P.
PR 05-SEP-2000; 2000US-000000P.
PR 18-SEP-2000; 2000US-0664610.
PR 18-SEP-2000; 2000US-0665350.
PR 24-OCT-2000; 2000US-242322P.
PR 08-NOV-2000; 2000US-0709238.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 22-JAN-2001; 2001US-0767609.
PR 28-FEB-2001; 2001US-0796498.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 25-MAY-2001; 2001WO-US17092.
PR 30-MAY-2001; 2001US-0870574.
PR 30-MAY-2001; 2001WO-US17443.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 28-JUN-2001; 2001WO-US00000.
XX (GETH) GENENTECH INC.
PA (BAKE/) BAKER K P.
PA (FERR/) FERRARA N.
PA (FERR/) FERRARA N.
PA (GERB/) GERBER H.
PA (GERB/) GERBTSSEN M E.
PA (GOD/) GODDARD A.
PA (GOD/) GODOWSKI P J.
PA (GURN/) GURNEY A L.
PA (HILL/) HILLAN K J.
PA (MARS/) MARSTERS S A.
PA (PANJ/) PAN J.
PA (PAON/) PAONI N F.
PA (STEP/) STEPHAN J F.
PA (WATA/) WATANABE C K.
PA (WILL/) WILLIAMS P M.
PA (WOOD/) WOOD W I.

XX Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
XX WPI; 2002-171999/22.
DR P-PSDB; ABB95436.
XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or angiogenic disorders in a mammal -
XX Claim 1; Fig 27; 567pp; English.

XX The present invention provides the protein and coding sequences of human
CC PRO proteins. These are useful for treating or diagnosing a
CC cardiovascular, endothelial or angiogenic disorder, including cardiac
CC hypertrophy, trauma, cancer, age-related macular degeneration,
CC atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound
CC healing. The present sequence is a coding sequence of the invention.

XX Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;

Alignment Scores:

| Pred. No.: | 2.57e-58 | Length: | 2005 |
|------------------------|----------|---------------|------|
| Score: | 1934.00 | Matches: | 341 |
| Percent Similarity: | 99.71% | Conservative: | 0 |
| Best Local Similarity: | 99.71% | Mismatches: | 1 |
| Query Match: | 99.69% | Indels: | 0 |
| DB: | 24 | Gaps: | 0 |

US-09-763-712a-2_COPY_206_547 (1-342) x ABL95574 (1-2005)

| | | | |
|----|------|---|------|
| Qy | 1 | MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle | 20 |
| Db | 478 | ATGCAACAAGATTGATGAGGTGCGAGGTAGACACTGAAGTAGCCCAACTATCAGTGATT | 537 |
| Qy | 21 | MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLeuLysAsnPheThrIle | 40 |
| Db | 538 | ATGGAAGAATGAAGCTAGTAGCTCCAAAGCATGGTCAGCTCATCAAGAATTTTACAATA | 597 |
| Qy | 41 | LeuGlnGlyProGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro | 60 |
| Db | 598 | CTACAAGGTCCACCGGGCCCCAGGGTCCAGAGGTGACAGAGGATCCCAGGGACCCCT | 657 |
| Qy | 61 | GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro | 80 |
| Db | 658 | GGCCCAACTGGCAACAGGACAGAAAGAGAGAGAGGGGAGCCTGGACACCCTGGCCCT | 717 |
| Qy | 81 | AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly | 100 |
| Db | 718 | GCGGGTGAGAGAGGCCCAATTGGACCAGCTGTCCCGCCCGAGAGCGTGGCGCAAGGA | 777 |
| Qy | 101 | SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln | 120 |
| Db | 778 | TCTAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAG | 837 |
| Qy | 121 | GlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyPro | 140 |
| Db | 838 | GGCCCCAGTGGGACCCCGCCCCCGGGCCCCACAGGCAAGAGGGGACTCCCGGCCCT | 897 |
| Qy | 141 | GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly | 160 |
| Db | 898 | CAGGGCCCTCCTGGGTTCAGGGGACTTCCAGGGCACCCTTGGGGAGCGTGGGGTGGCTGA | 957 |
| Qy | 161 | ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro | 180 |
| Db | 958 | CCTCGGGGACTGCGAGGCTTGCTGGGGTACCAGGATGCCAGGCCCGCCAGGGCCCCCCC | 1017 |
| Qy | 181 | GlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro | 200 |
| Db | 1018 | GGCCCTCCTGGCCCATCAGGAGGGGTGGTGGCCCTGGCCCTGCAGAAATGAGCAACCCG | 1077 |
| Qy | 201 | AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTrpTyr | 220 |
| Db | 1078 | GCACCGGAGGACAATAGCTGCCCGCTCAGTGGAGAACTTCACAGACAAATGCTACTAT | 1137 |
| Qy | 221 | PheSerValGlyLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer | 240 |
| Db | 1138 | TTTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAACAGACTTCCA | 1197 |
| Qy | 241 | HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly | 260 |
| Db | 1198 | CATCTGTGTTTCTAATAACACTAGAGAGGAGCAATGGATAAAAAACAGATGGTAGGG | 1257 |
| Qy | 261 | ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu | 280 |
| Db | 1258 | ACAGAGAGCCACTGGATCGGCTCACAGACTCAGACGCGTGAATAATGAATGAAGTGGCTG | 1317 |
| Qy | 281 | AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis | 300 |
| Db | 1318 | GATGGGACATCTCCAGACTACAAAAATTTGAAAGGTGGACAGCCGGATAAATCGGGTCA | 1377 |
| Qy | 301 | GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe | 320 |
| Db | 1378 | GGCCATGGCCAGAGAGAGACTGTGCTGGTTGATTATGCTGGCAGTGGAGACGATTTC | 1437 |

Db 958 CCTCGGGGACTCGCAGGCTTGCCTGGGTACCAAGGATGCCAGGCCCAAGGCCCCCC 1017
QY 181 GlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro 200
Db 1018 GGCCTCTCTGGCCATCAGGAGGGGTGGTCCCTGGCCTGCAGAAAGAGCAACCCG 1077
QY 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPhetThrAspLysCysTrpTyr 220
Db 1078 GCACCGGAGACAATAGTGCCTGCTCACTGGAAGAACTTCACACACAATGCTACTAT 1137
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSer 240
Db 1138 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGGAAGCAAGCTTCA 1197
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1198 CATCTTGTTTCATAAACAACACTAGAGAGAACACCAATGATGATAAACAAGATGGTAGG 1257
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 1258 AGAGAGAGCCACTGGATCGCCTCACAGACTCAGAGCGTGAAAAATGAATGGAAGTGGCTG 1317
QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 1318 GATGGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCGATAACTGGGGTCAT 1377
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 1378 GGCCATGGCCAGGACAGACTGTGCTGGTGTGATTTATGCTGGGCGAGTGGACGATTTC 1437
QY 321 GlnCysGluAspValAsnAsnPhelIleCysGluLysAspArgGluThrValLeuSerSer 340
Db 1438 CAATGTGAAGACGTCAATAACTTCAATTTGCGAAAAAGACAGGGAGACAGTACTGTCTATCT 1497
QY 341 AlaLeu 342
Db 1498 GCATTA 1503
RESULT 7
AAI60628
ID AAI60628 standard; cDNA; 2181 BP.
XX AAI60628;
XX XX
DT 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 4617.
XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX

(HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR P-PSDB; AAM41472.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PT
XX Claim 1: SEQ ID NO 4617; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2181 BP; 664 A; 539 C; 535 G; 443 T; 0 other;
Alignment Scores:
Pred. No.: 2,72e-58 Length: 2181
Score: 1934.00 Matches: 341
Percent Similarity: 99.71% Conservative: 0
Best Local Similarity: 99.71% Mismatches: 1
Query Match: 99.69% Indels: 0
DB: 22 Gaps: 0
US-09-763-712A-2_COPY_206_547 (1-342) x AAI60628 (1-2181)
QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 662 ATGCAACAAGATTGTGATGAGTTCGAGGTAGACACTGAAGTAGCAACTATCAGTGATT 721
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
Db 722 ATGGAAGAAATGAAGCTAGTAGACTCCAAGCATGGTCAAGCAATTTTACAATA 781
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
Db 782 CTACAGAGTCCACCGGGCCCCAGGGGTCCAAGAGGTGACAGAGATCCACGAGACCCCT 841
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGlyGluLysGlyGluProGlyProGlyPro 80
Db 842 GGCCCAACTGGCAACAAGGACAGAAAGGAGAGAGGGGAGCGCTGACACCTGGCCCT 901
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly 100
Db 902 GCGGGTGAGAGAGGCCCAATTGGACCAAGCTGGTCCCGCCGAGAGAGCGTGGCGCAAGGA 961
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
Db 962 TCTAAGAGCTCCAGAGGCCCAAGAGCTCCCGTGGTTCCTCGGAGAGCCCGCCCTCAG 1021
QY 121 GlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyPro 140
Db 1022 GGCCCCAGTGGGACCCAGGCCCGCCCCCGGCCACCAGGCAAGAGGAGCTCCCGGCCCT 1081
QY 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
|||||

Db 1082 CAGGGCCCTCCCTGGCTCCAGGACTTCAGGCACACCTGTGGGAGCCTGGGTGCTCGA 1141
 QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
 Db 1142 CTCGGGGACTCCAGGCTGTCTGGGTACAGCATGCCAGGCCCAAGGGCCCCCCC 1201
 QY 181 GlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro 200
 Db 1202 GGCCCTCCTGGCCATCAGGACGGTGTGCTCCCTGGCCCTCAGCAATGAGCAACCCG 1261
 QY 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnProHisTrpLysCysTrpTyr 220
 Db 1262 GCACCGGAGCAATAGCTGCCCGCTCCTCCTGGAAGAACTTCACAGCAAAATGCTACTAT 1321
 QY 221 PheSerValGluLysGluLeuPheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
 Db 1322 TTTTCAGTTGAGAAGAAATTTTTCAGGATGCAAGCTTTCTGTGAACACAGTCTTCA 1381
 QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
 Db 1382 CATCTTGTGTTTATAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGCTAGGG 1441
 QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
 Db 1442 AGAGAGCCACTGGATCGGCTCAGACTCAGAGCGTGAATGGAATGGAAGTGGCTG 1501
 QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
 Db 1502 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGGATAACTTGGGTCTAT 1561
 QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
 Db 1562 GGCCATGGGCCAGGAGAGACTGTCTGGGTGTGATTTATGCTGGCAGTGGAAACGATTTC 1621
 QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
 Db 1622 CAATGTGAGAGCTCAATAACTTCACTTCCGAAAAGACAGGGAGACAGTACTGTCTCT 1681
 QY 341 AlaLeu 342
 Db 1682 GCATTA 1687
 RESULT 8
 AAC66903
 ID AAC66903 standard; cDNA; 2641 BP.
 AC
 AC AAC66903;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Human EXMAD-14 coding sequence SEQ ID NO: 39.
 XX
 KW Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
 KW inflammation; reproductive disorder; cardiovascular disorder;
 KW immune disorder; musculoskeletal disorder; developmental disorder;
 KW gastrointestinal disorder; cell proliferation disorder; ss.
 XX
 OS Homo sapiens.
 PN WO200068380-A2.
 XX
 PD 16-NOV-2000.
 XX
 PF 10-MAY-2000; 2000WO-US12811.
 XX
 PR 11-MAY-1999; 99US-0133643.
 XX
 PR 23-AUG-1999; 99US-0150409.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DAM;
 PI Azimzal Y;
 XX

DR WPI; 2001-007395/01.
 DR P-PSDB; AAB27236.

XX Isolated polynucleotide encoding extracellular matrix or
 PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
 PT preventing disorders associated with expression of EXMAD such as
 PT proliferative, immune and genetic disorders -
 XX
 PS Claim 4; Page 121-122; 129pp; English.

XX The present invention provides the protein and coding sequences for 25
 CC novel extracellular matrix and adhesion-associated proteins (EXMADS).
 CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
 CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
 CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
 CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
 CC useful in the prevention and treatment of cancers, cell proliferation,
 CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
 CC gastrointestinal disorders and inflammation.

XX Sequence 2641 BP; 823 A; 628 C; 608 G; 582 T; 0 other;

Alignment Scores:

Pred. No.: 3,11e-58 Length: 2641
 Score: 1934.00 Matches: 341
 Percent Similarity: 99.71% Conservatives: 0
 Best Local Similarity: 99.71% Mismatches: 1
 Query Match: 99.69% Indels: 0
 DB: 0 Gaps: 0

US-09-763-712A-2_COPY_206_547 (1-342) x AAC66903 (1-2641)

QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspTrpGluValAlaAsnLeuSerValIle 20
 Db 912 ATGCACCAAGATTTGATGAGTTCGAGGTAGACACTGAAGTACCACTATATCATGAT 971
 QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLysAsnPhethrile 40
 Db 972 ATGGAAGAAATGAAGCTAGTACTCAAGCATGGTGCAGCTCATCAAGAAATTTTACAATA 1031
 QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
 Db 1032 CTACAAGGTTCACCGGGCCCCAGGGGTCCAGAGGTGACAGAGGATCCCGAGGACCCCT 1091
 QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
 Db 1092 GGCCCACTGGCAACAGGGACAGAAAGAGAGAGAGGGGAGCCTGGACCCTGGCCCT 1151
 QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly 100
 Db 1152 GCGGGTGAGAGAGGCCCAATTTGGACAGCTGCTCCCGCGGAGAGCGTGGCGGCAAGGA 1211
 QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
 Db 1212 TCTAAGGCTCCAGGGCCCCAAAGGCTCCCGTGTTCCTCGGGAAGCCCGGCCCTCAG 1271
 QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlyGluLeuProGlyPro 140
 Db 1272 GGCCCGAGTGGGACCCAGGCCCGCCCGGCCCCACAGGCAAGAGAGGACTCCCGGCCCT 1331
 QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
 Db 1332 CAGGGCCCTCTCTGGCTTCCAGGACTTCAGGGCACCGTTGGGGAGCCTGGGTGCTGGA 1391
 QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
 Db 1392 CCTCGGGGACTGCCAGGCTTGCCTGGGTACAGGATGCCAGGCTCCAGGCCCCAAGGCC 1451
 QY 181 GlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
 Db 1452 GGCCCTCTCTGGGCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAGTACAGCAACCCG 1511
 QY 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPhetThrAspLysCysTrpTyr 220

Db 1512 GCACCGGAGCAATAGCTGCCGCTCAGTGAAGAACTTACAGACAAATGCTACTAT 1571
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
Db 1572 TTTTCAGTTGAGAAAGAAATTTTGGAGATGCAAGCTTTTCTGTGAACAGACTTCA 1631
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1632 CATCTGTTTTCATAAACACTAGAGAGGAACAGCAATGATGATAAAAAACAGATGTTAGG 1691
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 1692 AGAGAGAGCCACTGGATCGGCTCACAGACTCAGACGCTGGAATGAATGAATGGAGTGGCTG 1751
QY 281 AspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 1752 GATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGACGCCGATAACTGGGGTCAAT 1811
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 1812 GCCATGGGCCAGGAGAACTGCTGGTGTGATTTATGCTGGCAGTGGAACTGTTTC 1871
QY 321 GlnCysGluAspValAsnAsnPhelleCysGluLysAspArgGluThrValLeuSerSer 340
Db 1872 CAATGTGAAGAGCTCAATAACTTCATTTTCGAAAAAGACAGGAGACAGTACTGTCTATCT 1931
QY 341 AlaLeu 342
Db 1932 GCATTA 1937

RESULT 9

ABQ92072

ID ABQ92072 standard; cDNA; 2929 BP.

XX ABQ92072;

XX 04-OCT-2002 (first entry)

XX Human polynucleotide SEQ ID NO 69.

DE Human; cytostatic; antirheumatic; antiarthritic; vulnery; analgesic;
KW antinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
KW neuroprotective; nootropic; osteopathic; haemostatic; vasotropic;
KW antitumor; fungicide; antidiabetic; antiasthmatic; antiallergic;
KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
KW stem cell; growth factor; nervous system disease; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
KW osteoporosis; severe combined immunodeficiency; SCID; infection;
KW multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss.

OS Homo sapiens.

XX US2002065394-A1.

XX 30-MAY-2002.

XX 22-DEC-2000; 2000US-0745763.

PR 18-MAR-1998; 98US-0040963.

XX (JACO/) JACOBS K.

PA (MCCO/) MCCOY J M.

PA (LAVA/) LAVALLIE E R.

PA (COLL/) COLLINS-RACIE L A.

PA (EVAN/) EVANS C.

PA (MERB/) MERBERG D.

PA (TREA/) TREACY M.

PA (SPAU/) SPAULDING V.

PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Spaulding V;

XX WPI; 2002-582343/62.
DR P-PSDB; ABP61859.
XX Novel secreted or transmembrane protein and polynucleotide encoding the
PT protein, useful for diagnosis and treatment of neurological disorders,
PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis
PT -
XX Claim 250; Page 231-232; 284pp; English.
XX The invention relates to human secreted or transmembrane protein (I),
CC their fragments and is encoded by specific complementary deoxyribonucleic
CC acid (cDNA) inserts (II), where the protein is substantially free from
CC other mammalian proteins. (I) are useful for preventing, treating or
CC ameliorating a medical condition, especially immunological treatment or
CC prevention of tumors. (I) exhibits activity relating to angiogenesis,
CC cytokine, cell proliferation, cell differentiation, antiinflammatory,
CC stem cell growth factor activity and activin or inhibin-related
CC rise to neuroepithelial cells that can be used to augment or replace
CC cells damaged by illness, autoimmune disease, accidental damage or
CC genetic disorders. (I) induces the proliferation of neural cells and
CC regeneration of nerve and brain tissue and is useful for the treatment of
CC central and peripheral nervous system diseases and neuropathies, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis. (I) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopaenia
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
CC periodontal disease. (I) is also useful for gut protection or
CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
CC in various tissues, various immune deficiencies and disorders including
CC severe combined immunodeficiency (SCID), bacterial or fungal infections,
CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
CC such as asthma or other respiratory problems. (II) is useful to express
CC recombinant protein, as markers for tissues in which the corresponding
CC protein is preferentially expressed and in gene therapy. The present
CC sequence is that of a polynucleotide of the invention.

XX Sequence 2929 BP; 901 A; 699 C; 692 G; 637 T; 0 other;

Alignment Scores:
Pred. No.: 3 33e-58 Length: 2929
Score: 1934.00 Matches: 341
Percent Similarity: 99.71% Conservative: 0
Best Local Similarity: 99.71% Mismatches: 1
Query Match: 99.69% Indels: 0
DB: 24 Gaps: 0

US-09-763-712a-2_COPY_206_547 (1-342) x ABQ92072 (1-2929)

QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1266 ATGCAACAAGATTTCATGATGAGTTCGAGGTTAGACACTGGAAGTAGCCAACTATCAGTGATT 1325
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhetHrIle 40
Db 1326 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGTGTACATCAAGAATTTTACAATA 1385
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
Db 1386 CTACAAGGTCCACCGGGCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCCGGACCCCT 1445
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 80
Db 1446 GCGCCCAACTGCGCAACAAGGACAGAAAGGAGAGAGGGGGGCGCTGGACACCTGGCCCT 1505
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlyGlyGlyGlyGlyGlyGlyGly 100

```

Db 1506 GCGGTGAGAGGCCCCAAATTTGGACCACTGGTCCCCCGGAGAGCGTGGCGGCAAAAGGA 1565
Qy 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
Db 1566 TCTAAGGCTCCAGGGCCCCAAAGGCTCCGCTGGTTCCTGGGAGCCCGGCGCTCAG 1625
Qy 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyProGlyLysGlyLeuProGlyPro 140
Db 1626 GGCCCCAGTGGGACCCAGGCGCCCGGCGCCACAGGCAAGAGGAGTCTCCCGGCGCT 1685
Qy 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
Db 1686 CAGGCGCTCTGCTCCAGGGACTTCAGGGACCCGTTGGGAGCCCTGGGCTGCTGGA 1745
Qy 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 1746 CCTGGGACTGCCAGGCTTGCCTGGGTACAGGATCCAGGCCCAAGGGCCCCCCC 1805
Qy 181 GlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
Db 1806 GGCCCTCTGCGCCATCAGGAGCGGTGGTCCCTGCGCTGCAGATGAGCAACCCG 1865
Qy 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyr 220
Db 1866 GCACGGAGGACAATAGCTGCCCGCTCCTCAGTGAAGAACTTCACAGACAAAATGCTACTAT 1925
Qy 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
Db 1926 TTTTCAGTTGAGAAGAATTTTGAGGATGCAAGCTTTCTGTGAAGACAAAGTCTCA 1985
Qy 241 HisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1986 CATCTTGTTCATAAACACATAGAGAGGACAGCAATGGATAAAAAACAGATGGTAGGG 2045
Qy 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 2046 AGAGAGACCACTGATCGGCTTCACAGCTCAGAGCTGAGAAATGAATGGAAGTGGCTG 2105
Qy 281 AspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 2106 GATGGACATCTCCAGACTACAAAATTTGGAAGCTGGACAGCCGAGTAACATGGGTCAT 2165
Qy 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 2166 GGCCATGGCCAGGAGAGACTGTCTGGTGTGATTATGCTGGCAGTGGAAACGATTTC 2225
Qy 321 GlnCysGluAspValAsnAspPheIleCysGluLysAspArgGluThrValLeuSerSer 340
Db 2226 CAATGTGAAGACGTCATTAACCTCATTTTGCAGAAAAGACAGGAGACAGTACTGTCTCT 2285
Qy 341 AlaLeu 342
Db 2286 GCATTA 2291
RESULT 10
AAV55746
ID AAV55746 standard; cDNA: 2930 BP.
XX
AC AAV55746;
XX
XX 23-MAR-1999 (first entry)
XX
XX Human secreted protein clone bv227_1 coding sequence.
XX
XX Secreted protein; human; nutritional supplements; immune stimulant;
KW immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis;
KW activin/inhibin; chemokinesis; haemostasis; thrombolysis;
KW receptor/ligand activity; anti-inflammatory; tumour inhibitor;
KW cadherin/tumour invasion suppressor; ds.
XX
OS Homo sapiens.
XX
XX key Location/Qualifiers
FH

```

```

FT CDS 67..693 /*tag= a
FT FT
XX XX
PN WO9855614-A2.
XX
PD 10-DEC-1998.
XX
PF 01-JUN-1998; 98WO-US11210.
XX
XX 29-MAY-1998; 98US-0087255.
PR 04-JUN-1997; 97US-0868696.
PR 04-JUN-1997; 97US-0868697.
PR 04-JUN-1997; 97US-0868698.
PR 04-JUN-1997; 97US-0868898.
PR 04-JUN-1997; 97US-0868899.
PR 04-JUN-1997; 97US-0868900.
PR 04-JUN-1997; 97US-0869191.
PR 04-JUN-1997; 97US-0869192.
PR 04-JUN-1997; 97US-0869193.
PR 04-JUN-1997; 97US-0869194.
XX
XX (GEMY ) GENETICS INST INC.
XX
XX Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER;
PI McCoy JM, Racie LA, Spaulding V, Treacy M;
XX
XX WPI: 1999-059912/05.
DR P-PSDB; AAW73628.
XX
XX New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult brain, foetal kidney, adult ovary, adult
PT retina, adult placenta or adult uterus cDNA libraries
XX
XX Claim 26; Page 87-88; 127pp; English.
XX
XX This sequence encodes a human secreted protein of the invention.
CC This DNA sequence was isolated from a human adult brain cDNA
CC library, and was designated clone bv227_1. The DNAs and proteins
CC are predicted to have biological activities which would make them
CC suitable for treating, preventing or ameliorating medical conditions in
CC humans and animals, although no supporting data is given. Suggested
CC activities include nutritional sources or supplements, immune
CC stimulating or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity. The DNAs
CC are also stated to be useful for gene therapy. A host cell transfected
CC with the DNA, or its subfragments and variants is useful for recombinant
CC production of the human secreted protein clones.
XX
SQ Sequence 2930 BP; 902 A; 699 C; 692 G; 637 T; 0 other;

Alignment Scores:
Pred. NO.: 3.33e-58 Length: 2930
Score: 1934.00 Matches: 341
Percent Similarity: 99.71% Conservative: 0
Best Local Similarity: 99.71% Mismatches: 1
Query Match: 99.69% Indels: 0
DB: 20 Gaps: 0

US-09-763-712A-2_COPY_206_547 (1-342) x AAV55746 (1-2930)
Qy 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1266 ATGCAACAGATTTTCAGGTTCGAGCTGACACTGAGTACAGTACCACTTATCAGTGAT 1325
Qy 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
Db 1326 ATGGGAAGAAATGAACCTAGTAGACTCCAGCATGCTCAGCTCATCAAGAAATTTACAATA 1385
Qy 41 LeuGlnGlyProGlyProArgGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60

```

```
Db 1386 CTACAAAGGTCCACCGGGCCCCAGGGGTCCAAAGAGGTGACAGAGGATCCACGGAGCCCCCT 1445
Qy 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
Db 1446 GCGCCAACTGGCAACAGGACAGAAAGAGAGAGAGGGGGAGCCCTGGACACCTGGCCCT 1505
Qy 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProAlaGlyProGlyGluArgGlyGlyLysGly 100
Db 1506 GCGGGTGAGAGAGGCCCAATTGGACAGCTGCTGCCCGGAGAGCGTGGCGCAAGGA 1565
Qy 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
Db 1566 TCTAAAGGCTCCAGGGCCCCAAAGGCTCCCGTGGTTCCTCCCTGGGAAGCCCGCCCTCAG 1625
Qy 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyPro 140
Db 1626 GCGCCAGTGGGACCCAGGCCCGCCCGGCCACACAGGCAAGAGGAGCTCCCGGCCCT 1685
Qy 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
Db 1686 CAGGGCCCTCCTGGCTTCCAGGGACTTCCAGGCACCGTTGGGGAGCCTGGGTGCGTGA 1745
Qy 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 1746 CCTCGGGGACTGCCAGGCTTGCTGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCCC 1805
Qy 181 GlyProProGlyProSerGlyAlaValProLeuAlaGluAsnGluProThrPro 200
Db 1806 GGCCTCTCTGGCCATCAGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGAGCCACCCG 1865
Qy 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTyr 220
Db 1866 GCACCGAGGACAATAGCTGCCCGCTCCTACTGGAAGAACTTCCAGACAAATGCTACTAT 1925
Qy 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSer 240
Db 1926 TTTTCAGTTGAGAAGAAATTTTGGAGATGCAAGCTTTCTGTAAGACAAGTCTTCA 1985
Qy 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1986 CATCTGTTTTCATAAACACTAGAGAGGAACAGCAATGATGATAAAAAACAGATGTTAGG 2045
Qy 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 2046 AGAGAGAGCCATGGATCGCCCTCAGACTCAGAGCGTGAATAATGAATGGAAGTGGCTG 2105
Qy 281 AspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 2106 GATGGGACATCTCCAGACTACAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCAT 2165
Qy 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 2166 GGCATGGGCCAGGAGAGACTGTGCTGGTGTATTTATGCTGGGAGTGAACGATTTTC 2225
Qy 321 GlnCysGluAspValAsnAsnPheIleCysGlyLysAspArgGluThrValLeuSerSer 340
Db 2226 CAATGTGAAGAGCTCAATAAATTCATTTCCGAAAAAGACAGGGAGACAGTACTGTCTATCT 2285
Qy 341 AlaLeu 342
Db 2286 GCATTA 2291
RESULT 11
AAH43037
ID AAH43037 standard; cDNA; 2637 BP.
XX
AC AAH43037;
XX
XX
DT 15-OCT-2001 (first entry)
DE Nucleotide sequence of a human scavenger receptor.
XX
KW Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;
```

```
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
KW low density lipoprotein; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 92..2320
FT a
FT /*tag=
FT /product= "scavenger receptor"
XX
XX WO200159107-A1.
XX 16-AUG-2001.
XX 08-FEB-2001; 2001WO-JP00874.
XX 14-FEB-2000; 2000JP-0035155.
XX 10-OCT-2000; 2000JP-0309068.
XX (FUSO ) FUSO PHARM IND LTD.
XX
XX Wakamiya N;
XX
XX WPI; 2001-497076/54.
XX P-P5DB; AAG63347.
XX
XX New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
XX useful for treatment and diagnosis of diseases associated with oxidized
XX low-density lipoprotein accumulation
XX
XX Claim 6; Page 88-93; 118pp; Japanese.
XX
XX The present sequence encodes a human scavenger receptor, designated
XX SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
XX are useful in clarifying the functions of macrophages and basal
XX immunity. They are also useful in the treatment, prevention, diagnosis
XX and investigation of diseases such as arteriosclerosis, diabetic
XX complications, bacterial infection and restenosis following angioplasty,
XX which are associated with accumulation of oxidized low density
XX lipoprotein and the binding of advanced glycation end-products into
XX cells.
XX
XX Sequence 2637 BP; 829 A; 617 C; 635 G; 556 T; 0 other;
```

```
Alignment Scores:
Pred. No.: 1.04e-53 Length: 2637
Score: 1801.00 Matches: 313
Percent Similarity: 95.91% Conservative: 15
Best Local Similarity: 91.52% Mismatches: 14
Query Match: 92.84% Indels: 0
DB: 22 Gaps: 0
US-09-763-712A-2_COPY_206_547 (1-342) x AAH43037 (1-2637)
Qy 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1292 ATGCGACAGACATGATGAGGTCAAGTTAGACACTGAAGTGGCCCAACTTATCACTGGT 1351
Qy 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuLysAsnPheThrIle 40
Db 1352 ATGGAGAGATGAACACTGGTGTACTCCCAAGCAGGTCACCTCATCAAGAACTTACCAT 1411
Qy 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
Db 1412 CTACAAGGTCTCTCTGGCCCCAGAGGTCCAAAAGGTGACAGAGGATCTCAGGACCACT 1471
Qy 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProPro 80
Db 1472 GGTCCAACCTGGCAACAAGGGACAGAAAGAGAGAGAGGGAGAGCCCTGGTCCACTGGCCCT 1531
Qy 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly 100
Db 1532 GCGGGTGAGAGGGGCACAAATTTGACCACCTGGCCCTCTCTGGAGAGCGGTGGCAGCAAGGA 1591
```



```
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
    |||
Db 1592 TCCAAAGGCTCACAGGTCCTCCAAAGATCTCTGGTCCCGAGGAAGCTGGCCCTCAA 1651
QY 121 GlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyPro 140
    |||
Db 1652 GGACCTAGTGGGGACCCAGGACCAAGCTCCAGGCAAGGATGGACTCCCTGGCCCT 1711
QY 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlyProGlyValProGly 160
    |||
Db 1712 CAGGGCCCTCCCTCCAGGAGCTACAGGCACTGTGGTGAGGCTCGAGTACCTGGA 1771
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
    |||
Db 1772 CCTCGGGGTTGCCAGGCTTCCAGGGTTCAGGGGTGCAGGCATCCCTTAAAGGACCACT 1831
QY 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
    |||
Db 1832 GGCCCTCCAGGCTCCCTCAGGACCAATTTGGAGCATTTGGCTCTCAGAAATGAACCAACCCCA 1891
QY 201 AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTrpTyr 220
    |||
Db 1892 GCATCAGAGCTCAACGAGTGTCCGCTCCTACTGGAAGAACTTCACAGATAAATGCTACTAT 1951
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
    |||
Db 1952 TTTTCATTGGAAGAAATTTTGAAGATGCTAAGCTTTCTGTGACACAAATCTTCC 2011
QY 241 HisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGly 260
    |||
Db 2012 CATCTCGTTTTCATAAATCAAGAGAAAGAACACGCAATGATGATAAAAGCATACCGTGGGG 2071
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
    |||
Db 2072 AGAGAAGCCATTGGATCGGCTCAGACTCAGAACAGAAAGCAATGGAAGTGGCTA 2131
QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
    |||
Db 2132 GACGGGTACCTGTTGATTACAAAACTGGAAGCTGGACCAACCATTAACCTGGGCGACT 2191
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
    |||
Db 2192 GGCCATGGCCAGGAGAACTGTCTGCTTCAATTTAGCGAGGACAGTGGAAATGACTTC 2251
QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
    |||
Db 2252 CAGTGTGATGAATCAATCAATCAATTCATTGTGTGAGAGGAAGGAGGAGTACCATCATCC 2311
QY 341 AlaLeu 342
    |||
Db 2312 ATATTA 2317
RESULT 12
AAS71133
ID AAS71133 standard; cdna; 1521 BP.
XX
AC AAS71133;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #6937.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
```

```
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI P-PSDB; ABG06946.
XX
DR WPI; 2001-639362/73.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 6937; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1521 BP; 386 A; 420 C; 425 G; 290 T; 0 other;

Alignment Scores:
Pred. No.: 5,8e-51 Length: 1521
Score: 1715.50 Matches: 298
Percent Similarity: 94.43% Conservative: 7
Best Local Similarity: 92.26% Mismatches: 7
Query Match: 88.43% Indels: 11
DB: 23 Gaps: 1

US-09-763-712A-2_COPY_206_547 (1-342) x AAS71133 (1-1521)
QY 17 LeuSerValIleMetGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLys 36
    |||
Db 130 CTACATGTCTCTCATTTATGCAATGAGAAATTTCTACCCGACAGG----- 171
QY 37 AsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySer 56
    |||
Db 172 -----CAAGGTCCACCGGGCCCGGCGGTCACAGAGAGAGAGAGAGAGAGATCC 216
QY 57 GlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGly 76
    |||
Db 217 CAGGGACCCCTGGCCCACTGGCAACAAGGGACAGAAAGGAGAGAGAGAGAGAGAGATCC 276
QY 77 ProGlyProProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArg 96
    |||
Db 277 CCACCTGGGCTCGGGGTGAGAGAGGCCCAATTTGGACCACTGTCCTCCCGGAGAGCGT 336
QY 97 GlyGlyLysSerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 116
    |||
Db 337 GCGGCAAGAGATCTAAGGCTCCAGGGCCCAAGAGCTCCCGTGGTTCCTCCCTGGGAG 396
QY 117 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyLysGluGly 136
```

```

Db 397 CCGGGCCCTCAGGGCTCCAGTGGGGACCCAGGCCCCCGGCCACAGGCAAGAGGGA 436
QY 137 LeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluPro 156
Db 457 CTCGCCGGCCCTCAGGGCCCTCTGGCTTCCAGGGACTTCAGGGCACCGTTGGGGAGCCT 516
QY 157 GlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyPro 176
Db 517 GGGGTGCTGGACTCGGGGACTGCCAGGCTTGCTGGGTACCCAGGATGCCAGGCCCC 576
QY 177 LysGlyProProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsn 196
Db 577 AAGGGCCCCCGGCCCTCTGGCCCATCAGGAGCGGTGGTGGCCCTGGCCCTGCGAGAAT 636
QY 197 GluProThrProAlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAsp 216
Db 637 GAGCAACCCACGACCGGAGGACATGGCTGCCGCCCTCAGTGGGAAGAACTTCACAGAC 696
QY 217 LysCysTyTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGlu 236
Db 697 AAATGCTACTATTTTTCAGTTGAGAAAGAAATTTTTCAGGATGCAAGCTTTTCTGTGAA 756
QY 237 AspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLys 256
Db 757 GACAAGCTCTTCACATCTTGTTCATAAACCACATAGAGAGGAACAGCAATGGATAAAAAA 816
QY 257 GlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGlu 276
Db 817 CAGATGGTAGGAGAGAGACCACTGGATCGGCTCCAGACTCAGAGCTCAGAGCGTGAAATGAA 876
QY 277 TrpLysTrpLeuAspGlyThrSerProAspTyTrpLysAsnTrpLysAlaGlyGlnProAsp 296
Db 877 TGGAACTGGCTGGATGGGACATCTCCAGACTACAAAAATTTGAAAGCTGGACAGCCGGAT 936
QY 297 AsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrrAlaGlyGln 316
Db 937 AACTGGGGTCATGGCCATGGGCGCAGAGAGAGACTGTGCTGGGTGTGATTTATGCTGGCGAG 996
QY 317 TrpAsnAspPheGlnCysGluAspValAsnAsnAsnPheIleCysGluLysAspArgGluThr 336
Db 997 TGGAAAGGATTTCAATGTGAAGACGTCAATACTTCATTTCGAAAAAGACAGGGAGACA 1056
QY 337 ValLeuSer 339
Db 1057 GTCTTCACG 1065

RESULT 13
AAH43054
ID AAH43054 standard; DNA; 2256 BP.
XX
AC AAH43054;
XX
XX
DT 15-OCT-2001 (first entry)
XX
DE Nucleotide sequence of a human scavenger receptor.
XX
KW Human; scavenger receptor; SRCL-PI; macrophage; basal immunity;
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
KW low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT 74..1936
FT /*tag= a
FT /product= "scavenger receptor"
XX
XX
PD WO200159107-A1.
XX
PD 16-AUG-2001.
XX
PD
XX
PF 08-FEB-2001; 2001WO-JP00874.

```

```

XX
PR 14-FEB-2000; 2000JP-0035155.
PR 10-OCT-2000; 2000JP-0309068.
XX
PA (FUSO ) FUSO PHARM IND LTD.
XX
PI Wakamiya N;
XX
XX WPI; 2001-497076/54.
DR P-PSDB; AAG63350.
XX
XX New Scavenger receptor proteins SRCL-PI with collectin-like structure,
PT useful for treatment and diagnosis of diseases associated with oxidized
PT low-density lipoprotein accumulation
XX
PS Claim 4; Page 105-109; 118pp; Japanese.
XX
XX The present sequence encodes a human scavenger receptor, designated
CC SRCL-PI. The SRCL-PI polypeptide has a collectin-like structure. They
CC are useful in clarifying the functions of macrophages and basal
CC immunity. They are also useful in the treatment, prevention, diagnosis
CC and investigation of diseases such as arteriosclerosis, diabetic
CC complications, bacterial infection and restenosis following angioplasty,
CC which are associated with accumulation of oxidized low density
CC lipoprotein and the binding of advanced glycation end-products into
CC cells.
XX
SQ Sequence 2256 BP; 721 A; 524 C; 528 G; 483 T; 0 other;

Alignment Scores:
Pred. No.: 7.12e-32 Length: 2256
Score: 1158.00 Matches: 218
Percent Similarity: 63.74% Conservative: 0
Best Local Similarity: 63.74% Mismatches: 0
Query Match: 59.69% Indels: 124
DB: 22 Gaps: 1

US-09-763-712A-2_COPY_206_547 (1-342) x AAH43054 (1-2256)
QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1274 ATGCAACACAGATTGATGAGTGCAGGTAGACACTGAAGTACCAACTATCAGTGATT 1333
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
Db 1334 ATGGAAGAAATGAAGCTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAATTTTACAATA 1393
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyProArgGlyAspArgGlySerGlnGlyPro 60
Db 1394 CTACAAGGTCCACCGGGCCCCAGGGTCCAAAGAGGTGACAGAGGATCCCAAGGACCCCT 1453
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyPro 80
Db 1454 GGGCCAACTGGCAACAGGGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1513
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyProGlyGluArgGlyGlyLysGly 100
Db 1514 GCG----- 1516
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
Db 1516 ----- 1516
QY 121 GlyProSerGlyAspProGlyProProGlyProProGlyProGlyLysGluGlyLeuProGlyPro 140
Db 1516 ----- 1516
QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
Db 1516 ----- 1516
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180

```

| | | | |
|---|--|--|------|
| PS | Claim 6; Page 142-150; 175pp; Japanese. | | |
| XX | This sequence encodes the human cellular stress response 3 (CSR3) protein of the invention. The CSR proteins are macrophage scavenger receptor proteins. The CSR proteins can be used in the treatment, gene therapy and diagnosis of diseases in which intracellular stress is important, such as arteriosclerosis, diabetic circulatory obstruction, and microbial infection. Expression of the proteins is induced in vivo in response to intracellular stress, and inhibits cell death as a result of such stress. | | |
| SQ | Sequence 3685 BP; 901 A; 1142 C; 972 G; 669 T; 1 other; | | |
| Alignment Scores: | | | |
| Pred. No.: | 1.07e-10 | Length: | 3685 |
| Score: | 540.00 | Matches: | 98 |
| Percent Similarity: | 66.14% | Conservative: | 27 |
| Best Local Similarity: | 51.85% | Mismatches: | 46 |
| Query Match: | 27.84% | Indels: | 18 |
| DB: | 20 | Gaps: | 2 |
| US-09-763-712A-2_COPY_206_547 (1-342) x AAX27858 (1-3685) | | | |
| QY | 5 | LeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMet | 24 |
| DB | 1405 | CTGCTCAGTCCCGCGTGGACCTCAACGCTCCGGAACCTCTCCATGATCTGGAGGAGATG | 1464 |
| QY | 25 | LysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyPro | 44 |
| DB | 1465 | AAGGCGATGGACACACAGCATGGAGAAATCTTCGCAATGTCCACATCTCAGAGGTGCC | 1524 |
| QY | 45 | ProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGly | 64 |
| DB | 1525 | CCCGGCCCTCCAGGACCAAGAGGATTCAAGAGGAGATATGGCGTGAAGGCGCTGTGGC | 1584 |
| QY | 65 | AsnLysGlyGlnLysGlyGlu-----Lys 72 | |
| DB | 1585 | GGCAGAGCCCGAAAGAGAGACCCCGGCATCTTTGGCCCTGGGACCCAGGCTCTCTCAG | 1644 |
| QY | 73 | GlyGluProGlyProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyPro | 92 |
| DB | 1645 | GGGCAACCTGGAGAGCGCGCTGTGGAGAAAGGGCGCTGTGGCCCTCGAGGGTTC | 1704 |
| QY | 93 | ProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGly | 112 |
| DB | 1705 | CCAGGCCTCAAAGGCTCAAAGGCGAGCTTTGGAACCTGGAGGCGCAG-----1752 | |
| QY | 113 | SerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyProProGlyProPro | 132 |
| DB | 1753 | -----GGACAGCCAGGCCCAAGGGGACATAGGCCCCCAGGGCCAGAGGCCCGC | 1806 |
| QY | 133 | GlyLysGluGlyLeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThr | 152 |
| DB | 1807 | GGGTCTCCAGGCGCTCAGGCGCTCAGGGAACCGGGAATTCAGGGGAGACAGGCTCA | 1866 |
| QY | 153 | ValGlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly | 172 |
| DB | 1867 | CCAGGCCAGCGGGGCCCATGGGCCCTAAGGTGAACACGAGGATCCAGGCTCCCGCTGT | 1926 |
| QY | 173 | MetProGlyProLysGlyProProGly 181 | |
| DB | 1927 | CTCCCGGGCGCTCCAGGTCACACAGGA 1953 | |
| RESULT | 15 | | |
| ID | AAAX27856 | | |
| XX | AAAX27856 standard; DNA; 3810 BP. | | |
| XX | AAAX27856; | | |
| XX | 02-JUN-1999 | (first entry) | |
| XX | Human CSR1 | protein coding sequence. | |
| XX | Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage; | | |

KW scavenger receptor protein; intracellular stress; arteriosclerosis;
KW diabetic circulatory obstruction; microbial infection; ss.
XX Homo sapiens.
OS WO9909159-A1.
PN 25-FEB-1999.
XX 12-AUG-1998; 98WO-JP03602.
XX 30-JUL-1998; 98JP-0230121.
PR 13-AUG-1997; 97JP-0233396.
XX (NISB) JAPAN TOBACCO INC.
PI Nakamura Y, Tokino T;
XX WPI; 1999-181032/15.
DR P-PSDB; AAY00992.
XX Scavenger receptor proteins - for treatment and diagnosis of
XX disorders involving cell stress
PS Claim 4; Page 119-127; 175pp; Japanese.
XX This sequence encodes the human cellular stress response 1 (CSR1) protein
CC of the invention. The CSR proteins are macrophage scavenger receptor
CC proteins. The CSR proteins can be used in the treatment, gene therapy
CC and diagnosis of diseases in which intracellular stress is important,
CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial
CC infection. Expression of the proteins is induced in vivo in response to
CC intracellular stress, and inhibits cell death as a result of such stress.
XX
SQ Sequence 3810 BP; 892 A; 1213 C; 1049 G; 655 T; 1 other;

Alignment Scores:

| | | | |
|------------------------|---------|---------------|------|
| Pred. No.: | 1.1e-10 | Length: | 3810 |
| Score: | 540.00 | Matches: | 98 |
| Percent Similarity: | 66.14% | Conservative: | 27 |
| Best Local Similarity: | 51.85% | Mismatches: | 46 |
| Query Match: | 27.84% | Indels: | 18 |
| DB: | 20 | Gaps: | 2 |

US-09-763-712a-2_COPY_206_547 (1-342) x AAX27856 (1-3810)

| | | | |
|----|------|--|------|
| QY | 5 | LeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMet | 24 |
| Db | 1530 | CTGCTCAGTGCCTCCGGCTGGACCTCAACGTCGGAACTCTCCATGATCGTGGAGGATG | 1589 |
| QY | 25 | LysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyPro | 44 |
| Db | 1590 | AAGGCAGTGGACACACAGCATGGAGAAATCCTTCGCAATGTCACCATCTACGAGGTGCC | 1649 |
| QY | 45 | ProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGly | 64 |
| Db | 1650 | CCCGGCCCTCCAGGACCAAGAGGATTCAAAGGAGATATGGCGCTGAAAGGGCTGTGGC | 1709 |
| QY | 65 | AsnLysGlyGlnLysGlyGlu-----Lys | 72 |
| Db | 1710 | GGCAGAGGCCCGAAAGGAGACCCCGCATCTTGGGGCCCTGGGACCCAGGGTCTCTCAG | 1769 |
| QY | 73 | GlyGluProGlyProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyPro | 92 |
| Db | 1770 | GGCAACCTGGAGAGCCCGGCTGTGGGAGAAAGGGCCCTGTGGCCCTCGAGGGTTC | 1829 |
| QY | 93 | ProGlyGluArgGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGly | 112 |
| Db | 1830 | CCAGGCCCTCAAAGGCTCAAAGGGCAGCTTTGGAACCTGGAGGGCCGAGA----- | 1877 |
| QY | 113 | SerProGlyLysProGlnGlyProSerGlyAspProGlyProProGlyProPro | 132 |
| Db | 1878 | -----GGACAGCCAGGCCCAAAAGGGGACATAGGCCCCCAGGGCCCAAGGGCCCCCG | 1931 |

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:26:21 ; Search time 36.3343 Seconds

(without alignments)

2886.622 Million cell updates/sec

Title: US-09-763-712A-2_COPY_206_547

Perfect score: 1940

Sequence: 1 MQQDLMSRLDTEVANLSVI.....EDVNNFICKDRETVLSSAL 342

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 1533381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09763712/runat_14032003_100951_18187/app_query.fasta_1.1877
-DB=Issued_Patents_NA -QWTF=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPECL=0
-LOOPEXT=0 -UNITS=dits -START=1 -END=1 -MATRIX=diosum62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712@cgn_1_108 @runat_14032003_100951_18187 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6CTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|------------------|
| 1 | 459.5 | 23.7 | 810 | 1 | US-08-642-255-60 |
| 2 | 452.5 | 23.3 | 3181 | 1 | US-08-655-086-1 |
| 3 | 448.5 | 23.1 | 756 | 1 | US-08-642-255-50 |
| 4 | 436.5 | 22.5 | 1608 | 4 | US-09-029-348-19 |
| 5 | 428 | 22.1 | 4031 | 1 | US-08-159-784-1 |
| 6 | 424.5 | 21.9 | 1572 | 4 | US-09-297-269-39 |
| 7 | 419 | 21.6 | 2543 | 1 | US-08-555-669-11 |
| 8 | 419 | 21.6 | 2543 | 3 | US-09-073-663-11 |
| 9 | 412.5 | 21.3 | 432 | 1 | US-08-642-255-48 |
| 10 | 409 | 21.1 | 3394 | 1 | US-08-159-784-4 |
| 11 | 402.5 | 20.7 | 1416 | 1 | US-07-621-091G-1 |
| 12 | 402.5 | 20.7 | 1416 | 2 | US-08-399-889-1 |

Sequence 1, Appli
Sequence 1, Appli
Sequence 4, Appli
Sequence 20, Appli
Sequence 5, Appli
Sequence 1, Appli
Sequence 40, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 264, App
Sequence 62, Appli
Sequence 9, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 372, App
Sequence 137, App
Sequence 94, Appli
Sequence 66, Appli
Sequence 116, App
Sequence 3, Appli
Sequence 1, Appli
Sequence 3, Appli
Sequence 4, Appli
Sequence 6, Appli
Sequence 1, Appli
Sequence 50, Appli

13 402.5 20.7 1416 3 US-09-167-364-1
14 402.5 20.7 1416 4 US-09-439-897-1
15 400.5 20.6 4359 4 US-09-484-970B-4
16 399 20.6 1881 4 US-09-029-348-20
17 394.5 20.3 1560 2 US-08-794-795-5
18 394.5 20.3 1560 4 US-09-249-200-5
19 394 20.3 5102 1 US-08-494-168-1
20 393.5 20.3 503 4 US-09-297-269-40
21 389.5 20.1 1703 2 US-08-794-795-1
22 389.5 20.1 1703 4 US-09-249-200-1
23 380.5 19.6 1868 1 US-08-392-367B-1
24 380.5 19.6 1868 3 US-08-893-467A-1
25 370 19.1 1560 4 US-09-453-702B-264
26 370 19.1 61663 4 US-09-453-702B-62
27 355.5 18.3 2383 4 US-09-523-487-9
28 355.5 18.3 2409 3 US-09-320-095-9
29 350 18.0 1458 4 US-09-111-470-3
30 344.5 17.8 1839 1 US-08-383-744-1
31 344.5 17.8 1839 2 US-08-999-336-1
32 344.5 17.8 1839 5 PCT-US96-01427-1
33 330.5 17.0 392 4 US-09-404-879A-372
34 325.5 16.8 48908 4 US-09-453-702B-137
35 323.5 16.7 3404 4 US-09-453-702B-94
36 322 16.6 1370 4 US-09-111-470-9
37 321.5 16.6 9827 4 US-09-453-702B-66
38 320.5 16.5 45175 4 US-09-453-702B-116
39 316.5 16.3 2824 2 US-09-010-928B-3
40 315.5 16.3 2851 4 US-09-535-521-1
41 315.5 16.3 2851 4 US-09-535-521-3
42 310.5 16.0 876 4 US-09-535-521-4
43 310.5 16.0 876 4 US-09-535-521-6
44 306 15.8 18609 4 US-08-943-731-1
45 305.5 15.7 756 1 US-08-642-255-50

ALIGNMENTS

RESULT 1
US-08-642-255-60
; Sequence 60, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:

```

:
: COUNTRY: US
: ZIP: 11553
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: Patent In Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
:

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/655,086
: FILING DATE: 03-JUN-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: STEEN, JEFFREY S.
: REGISTRATION NUMBER: 32,063
: REFERENCE/DOCKET NUMBER: 203-1632
: TELECOMMUNICATION INFORMATION:
:

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-228-8484

```

/ TELEFAX: 516-228-8516
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3181 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: unknown
/ MOLECULE TYPE: cDNA
/ US-08-655-086-1

```

Alignment Scores:
Pred. No.: 1.57e-16 Length: 3181

| | | | |
|------------------------|--------|---------------|----|
| Score: | 452.50 | Matches: | 99 |
| Percent Similarity: | 52.11% | Conservative: | 12 |
| Best Local Similarity: | 46.48% | Mismatches: | 53 |
| Query Match: | 23.32% | Indels: | 49 |
| DB: | 1 | Gaps: | 8 |

| DB: | I | Gaps: | B |
|-----------------------|-----------------|-------------------|----------|
| US-09-763-712A-2 COPY | 206 547 (1-342) | x US-08-655-086-1 | (1-3181) |

| | | | |
|--|----|---|-----|
| | Qy | LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlnPro | 60 |
| | | :::::::::::::::::::: | |
| | | :::::::::::::::::::: | |
| | Db | GTTCAAGACCCCTGGCCCTGCTGGAGGAAGAAGCGAGGACTCGAGTGAACCC | 952 |

Qy 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro----- 75

| | | |
|----|-----|---|
| DD | 953 | GGACCCACATGGCCCTTCCCTCCCGGACCCCCCTTGGCGAAGACCTGGTGGACCTTGGTAGCCCGCGGGTTTC |
| QY | 76 | -----GlyProProGlyProAlaGlyGluArgGlyProIleGly |

Db 1013 CCTGGCGCAGATGGTGTGCTGGTCCCCAGGGTCCCCGGTGGTGAACGTGGTTCTCTCTGGC 1072

[illegible]

Qy 106 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 125

| | | | |
|----|-----|--|-----|
| QY | 126 | ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProPro--- | 144 |
|----|-----|--|-----|

[illegible]

Qy 145 -----GlyPheGlnGlyLeuGlnGlyThrValGlyGluPro 156
Db 1244 GCCCGTGGTCAGCGTGGTGTGATGGGATTCCTCGACCTAAAGSGTCTGCTGGAGAGCC 1303

| | | | | |
|----|------|--|-------|------|
| Qv | 157 | glvvalproGlyproArgGlyLeuProGlyLeuProGlyval | ----- | 170 |
| Db | 1244 | GCCCGTGGTCAGGCGTGGTGTGATGGGATCCCTGGACCTAAAGGTGCTGCTGGAGAGCCC | | 1303 |

| | | | | |
|----|------|--|-------|------|
| Qy | 157 | GlyValProGlyProArgGlyLeuProGlyVal | ----- | 170 |
| | | | | |
| | | | | |
| | | | | |
| Db | 1304 | GGCAAGGCTGGAGACGAGGTGTTCCCGACCCCTGTCGGTCTCTGCTGGCAAA | | 1363 |

| | | | |
|----|------|---|-----|
| DB | 1304 | GGCAGGCTGGAGACGAGGTTCTCCCGGACCCCCCTGGCGCTGTCTGGTCCCTGCTGGCAAA | 136 |
| Qy | 171 | ProGlyMetProGlyProLysGlyProGlyProGlyProGlyProSerGly----- | 187 |

[illegible][illegible]

```

QY 188 -----AlaValValProLeuAlaLeuGlnAsnGluProThrProAla----- 201
Db 1424 GAACAGGCGCTGCTGGCTCCCGCC---GGATTCCAGGGTCTCCTGGTCTGCTGCTCT 1480
QY 202 -----ProGluAspAsnGlyCysPro 208
Db 1481 CCAGGTGAAGCAGGCAACCTGGTGAACAGGGTGTCTCT 1519

RESULT 3
US-08-642-255-50
; Sequence 50, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A5556-3/BI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 756 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic"
US-08-642-255-50

Alignment Scores:
Pred. No.: 6,41e-17 Length: 756
Score: 448.50 Matches: 87
Percent Similarity: 55.03% Conservative: 6
Best Local Similarity: 51.48% Mismatches: 63
Query Match: 23.12% Indels: 13
DB: 1 Gaps: 2

US-09-763-712a-2_COPY_206_547 (1-342) x US-08-642-255-50 (1-756)
QY 43 GlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 62
Db 10 GCGCCAGCAGGTCCGAGGCGCGGATGCCCGCAGGCGCCGAGGCGGCTGGACCG 69
QY 63 ThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGly 82
Db 70 GCTGGTCCACCGGGTCTCCGGGACCTGCGAGCCCGCCAGGTGGCGCTGGACCGGTGT 129
QY 83 GluArgGlyProIleGlyProAlaGlyProGlyProGlyGluArgGlyClyLysGlySerLys 102
Db 130 CCACCGGGTGTCTCCGGGACCTGCAGCGCCGCGAGGTGCGCTGGACCGGTGTCTCCACCG 189

```

```

QY 103 GlySerGlnGlyProLysGlySerArgGlySerPro-----GlyLysProGlyPro 119
Db 190 GGTGCTCCGGACCTGCAGCGCCCGCAGGTGGCGCTGGACCGGCTGTCCACCGGGTGCT 249
QY 120 GlnGlyProSerGlyAspProGlyProGlyProGlyProGlyProGlyLysGluGlyLeuProGly 139
Db 250 CCGGGACCTGCAGCGCCCGCAGGTGGCGCTGGACCGGCTGTCCACCGGGTGTCTCCGGA 309
QY 140 ProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValPro 159
Db 310 CCTGAGGCGCCGCGAGGTGGCGCTGGACCGGCTGTCCACCGGCTGTCCGGGACCTGCA 369
QY 160 GlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyPro 179
Db 370 GCGCCGCGAGGTGGCGCTGGACCGGCTGTCCACCGGCTGTCCGGGACCTGCAGGCGCG 429
QY 180 ProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThr 199
Db 430 CCAGGTGGCGCTGGACCGGTGT-----CCACCG 459
QY 200 ProAlaProGluAspAsnGlyCysPro 208
Db 460 GGTGCTCCGGGACCTGCAGCGCGCGCA 486

RESULT 4
US-09-029-348-19
; Sequence 19, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: d087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1608
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-19

Alignment Scores:
Pred. No.: 5,93e-16 Length: 1608
Score: 436.50 Matches: 102
Percent Similarity: 47.06% Conservative: 18
Best Local Similarity: 40.00% Mismatches: 67
Query Match: 22.50% Indels: 70
DB: 4 Gaps: 9

US-09-763-712a-2_COPY_206_547 (1-342) x US-09-029-348-19 (1-1608)
QY 30 LysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProGlyProArgGly 49
Db 59 CAACATGCCCAATCTTTACAGAGGAA-CTGTGAAGAGGCGCCCGGAGATAGAGGA 117
QY 50 ProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGlyAsnLysGlnLys 69
Db 118 CCACGTGGAGAAAGGGTCCACAGCGCCCGCAGGAGAGATGTTGAAGATGTCCACA 177
QY 70 GlyGluLysGlyGluProGlyProGlyPro----- 80
Db 178 GCGCCTCTCTGGTCCACCTGTCTCTGGCGCCCGCTCGTGGTGGTGGTGGTGGTGGT 237
QY 81 -----AlaGlyGluArgGlyProIleGlyProAla 90
Db 238 CAGTATATGGAAGAGAGTGGACCTGGCCCTGGACCAATGGGCTTAATGGGACCTAGA 297
QY 91 GlyProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySer 110

```

```

|||||
298 GGCCCACTGGTGCAGTCGAGCCCTCAAGGTTTCCAAGGACCTGCTGGTGA 357
QY 111 ArgGlySerProGlyLys -----ProGlyProGlnGlyPro----- 122
Db 358 CTGGTGAACCTGGTCAAACTGGTCTCGAGGTGCACCTGGTCTCATGGCCCGCTGGGT 417
QY 123 -----SerGlyAspProGlyProProGlyProProGlyLysGlu 135
Db 418 CTGCTGGCAAAACATGGAACCTGGTGAACCTGGTCTCTGCTGCTGCTGCTGCT 477
QY 136 GlyLeuProGlnGlyProGlnGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 155
Db 478 GGTGCTGTTGGCCCAAGAGGTCTAGTGGCCCAAGGATTCGTGGCGATAAGGAGAG 537
QY 156 ProGlyValProGlyProArgGlyLeuPro-----Gly 166
Db 538 CCGCGTGAAGGGGCCAGAGGTCTCTCGGCTTCAAGGGACACAATGATGTCGAAGT 597
QY 167 LeuProGlyVal -----ProGlyMetProGlyProLys 177
Db 598 CTGCTGGTATGCTGCTGCTCAACATGTGTCAAGGTGCTCTGCTGCTGCTGCTGCT 657
QY 178 GlyProGlyProGlyProGlyProSerGly----- 187
Db 658 GTCTCTAGGGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717
QY 188 ---AlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGly 206
Db 718 CTGTGTACGGTGGACCTGCTGG-CATTGAGGCCCTCAGGTGCACCAAGGCC- 770
QY 207 Cys---ProPro-----HisTrpLysAsnPhetheAspLysCys 218
Db 771 TGCTGGCCGCCCTGGTCCGCCCTCCTGGCCCTCCTGGACCTCCAGGTTGT 815
```

RESULT 5

```

US-08-159-784-1
; Sequence 1, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4031
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-159-784-1
Alignment Scores:
Pred. No.: 4,17e-15 Length: 4031
Score: 428.00 Matches: 95
Percent Similarity: 46.41% Conservative: 15
Best Local Similarity: 40.08% Mismatches: 49
Query Match: 22.06% Indels: 78
DB: 1 Gaps: 8
US-09-763-712A-2_COPY_206_547 (1-342) x US-08-159-784-1 (1-4031)
QY 26 LeuValAspSerLysHisGlyGlnLeuLeuLysAsnPhetheLeuGlnGlyProPro 45
Db 976 GTGGTCCAGAGCCCCCACTCAACACCTGTCCCTGGA-----GCACAAGGACCCCCG 1026
QY 46 GlyProArgGlyPro----- 50
Db 1027 GGACCTCAGGGGCCACCAGGGAAGGATGGCACTCCAGGAAGGATGGTGAACCGGTGAC 1086
QY 51 -----ArgGlyAspArgGlySerGlnGlyProProGlyProThrGly 64
Db 1087 CCTGTGAAGATGGGAGACCGGGTGCACCTGACCTCAAGGCTTTCAGGGACCCAGGA 1146
QY 65 AsnLysGlnLysGlyGluLysGlyGluPro-----GlyProProGly 79
Db 1147 GATGTGGGCCCTAAGGGCCGAGAAGGAGATCTCTGTGATTGGGGCCCCGAGACCTCCAGG 1206
QY 80 ProAlaGlyGluArgGlyProLe----- 87
Db 1207 CCTCCAGGGCCACCAGGACCCCTCTTCAGACAAGACAAGCTTCAATTGACATGGAG 1266
QY 88 -----GlyPro 89
Db 1267 GGATCCGGTTTCAGCGGAGACATAGAGGCTTTAGAGGCCACAGAGGCTTCCCTGGCCCC 1326
QY 90 AlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGly 103
Db 1327 CCGGGCCCCCTGGTGTCCAGGACCTTCTGTGTGACCGAGGCTTTGGATCAATGGT 1386
QY 104 SerGlnGlyProLysSerArgGlySerProGlyLysProGlnGlyProSer 123
Db 1387 TCCTATGCACCA---GGACCTGCAGGCTTCTCTGTGTACCTGGGAAGGAAGACCCCCC 1443
QY 124 GlyAspProGlyProProGlyProGlyLysGluGlyLeuProGlyProGlnGlyPro 143
Db 1444 GGTTCCTCAGGTCCCGGGACCTCCAGGTCCT-----CCAGGCAAGAGGGGCCCA 1494
QY 144 ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArg 162
Db 1495 CCAGGAGTGGCCGCCAGAAAGGCAGTGTGTGTGGCATCCAGGACCCAGGGG 1554
QY 163 -----GlyLeuProGlyLeuProGlyValProGlyMetPro 174
Db 1555 AGCAAGAGGACACCTTGGGCCCATCGGTATGCTTGGCAAGTCTGGCTGGTGGATCCCT 1614
QY 175 GlyProLysGlyProProGlyProProGlyProSerGlyAlaValProPro 191
Db 1615 GGGCCAGTTGGACCCCGGACCTCCAGGCGCTCCAGGGCCCTCCAGGACCA 1665
RESULT 6
US-09-297-269-39
; Sequence 39, Application US/09297269
; Patent No. 6451557
; GENERAL INFORMATION:
; APPLICANT: VAUGHAN, Paul R.
; APPLICANT: GALANIS, Maria
; APPLICANT: RAMSHAW, John A.M.
; APPLICANT: WERKMEISTER, Jerome A.
; TITLE OF INVENTION: A METHOD FOR PRODUCING, IN YEAST, A HYDROXYLATED TRIPLE
```


; TITLE OF INVENTION: HELICAL PROTEIN, AND YEAST HOST CELLS USEFUL IN SAID
; TITLE OF INVENTION: METHOD (As Amended)
; FILE REFERENCE: Q54094
; CURRENT APPLICATION NUMBER: US/09/297,269
; CURRENT FILING DATE: 1999-04-28
; EARLIER APPLICATION NUMBER: P03310
; EARLIER FILING DATE: 1996-10-29
; EARLIER APPLICATION NUMBER: P04306
; EARLIER FILING DATE: 1996-12-19
; EARLIER APPLICATION NUMBER: PCT/AU97/00721
; EARLIER FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 39
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Collagen Type III Alpha I Chain
US-09-297-269-39

Alignment Scores:
Pred. No.: 2,58e-15 Length: 1572
Score: 424.50 Matches: 119
Percent Similarity: 42.48% Conservativity: 25
Best Local Similarity: 35.10% Mismatches: 108
Query Match: 21.88% Indels: 88
DB: 4 Gaps: 15

US-09-763-712A-2_COPY_206_547 (1-342) x US-09-297-269-39 (1-1572)

Qy 19 ValIleMetGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhe 38
Db 199 :::::::::::::::::::: |||
Db 194 GTGAACCTGGGAAGAGATGGAACCTGGAT-----CAGATGGTC-----TTC 235
Qy 39 ThrIleLeuGlnGlyProGlyProArgGlyProArgGlyProArgGlyAspArgGlySer----- 56
Db 236 CAGGTGCGA-GATGGATCTCTGGTGGCAAGGATGATCGTGGTGAAGAAATGGCTCTCTCGT 294
Qy 57 -----GlnGlyProGlyProGlyProGlyProGlyProGlyProGlyGlnLys 69
Db 295 GCCCTGGCGCTCTGGTGCATCCAGGCCACCTGGTCTGCTGGTCCAGCTGGAAGAGT 354
Qy 70 GlyGluLysGlyGluProGlyPro-----ProGlyProAlaGlyGlu 83
Db 355 GGTGACAGAGGAGAAGTGGCCCTCTGGCCCTGCTGGTCTCCGGTCTCGCTGCTGCC 414
Qy 84 ArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGly 103
Db 415 CGAGGTGCTCTGGTCTCTCAAGGCCACCTGGTGTGCAAGAGGTGAACAGGTGAACGTGA 474
Qy 104 SerGlnGlyProLysGlySerArg-----GlySerProGlyLysPro 117
Db 475 GCTGTGTCATCAAGAGACATCGAGGATTCCTGTGTAATCCAGGTGCCAGGTTCCTCA 534
Qy 118 GlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlyGluLeu 137
Db 535 GGCCTGTGTCAGCAGGTGCAATCGGCAGTCCAGGACCTGCAGGCCCCAGAGGACCT 594
Qy 138 ProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGly 157
Db 595 GTTGGACCCAGTGGACCTCTGGCAAGATGGAACCATGGAGATCCAGGTCCCATTTGA 654
Qy 158 ValProGlyPro-----ArgGlyLeuProGlyLeuProGlyValPro 171
Db 655 CCACGAGGCTCGAGGTAAACAGAGTGAAGAGAGATCTGAGGGTCCCCAGGCCACCCA 714
Qy 172 GlyMetProGlyProLysGlyProGlyProGlyProGlyProSer-----Gly 187
Db 715 GGGCAACAGGCGCTCTCTGGACCTCTGTCGCCCTGCTGCTGCTGCTGCTGCTGCTGGA 774
Qy 188 AlaValProLeuAlaLeuGlnAsnGlu-----ProThrProAla 201
Db 775 GCCGCTGCCATTGCTGGGATTGGAGGTGAAAGAGTGGCGGTTTGTGCCCGCTATTATGGA 834

Qy 202 ProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAsp----- 216
Db 835 CTTGAA-----CCAATGGATTTCAAAAATCAACACCGATGAGATTACTTCA 882
Qy 217 ---LysCysTyrPheSerValGluLysGluIlePheGluAspAlaLys----- 232
Db 883 CTCAGTCTGTAAATGGACAATAAGAACCCATAGTCCTGATGGTTCTCGTAAAC 942
Qy 233 -----LeuPheCysGluAspLys----- 238
Db 943 CCCGCTAGAAACTGCAGAGACCTGAAATCTGCCACTCCTGAACCTCAAGACTGGAGAATAC 1002
Qy 239 -----SerSerHisLeuValPheIleAsnThrArg 248
Db 1003 TGGGTGACCCCTAACCAAGATGCAAAATGGATGATCAAGGTATCTGTATATGGAA 1062
Qy 249 GluGluGlnGlnTrpIleLysLysGlnMetValGly---ArgGluSerHisTrpIleGly 267
Db 1063 ACTGGGAAACATGCATAGTCCCAATCCTTTGAATGTTCCACGGAACACTGG----- 1116
Qy 268 LeuThrAspSerGluArgGluAsnGluTrpLysTrp-----LeuAspGly 282
Db 1117 TGGACAGATTCTAGTCTGAGAGAAGAACACACGTTTGGTTGGAGAGTCCATCGATGGT 1173
RESULT 7
US-08-555-669-11
; Sequence 11, Application US/08555669
; Patent No. 5773248
; GENERAL INFORMATION:
; APPLICANT: Brewton, Richard G.
; APPLICANT: Mayne, Richard
; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,669
; FILING DATE: 13-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: 8389-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-854-3660
; TELEFAX: 415-854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..2098
US-08-555-669-11

Alignment Scores:
Pred. No.: 8.17e-15 Length: 2543
Score: 419.00 Matches: 121
Percent Similarity: 35.59% Conservativity: 26

```

Best Local Similarity: 29.30% Mismatches: 91
Query Match: 21.60% Indels: 176
DB: 1 Gaps: 9

US-09-763-712a-2_COPY_206_547 (1-342) x US-08-555-669-11 (1-2543)

QY 27 ValAspSerLysHisGlyGlnLeuLeuLysAsnPheThrIleLeuGlnGlyProGly 46
DB 986 CTCGATGGCCAGAGGAGAGCTGGTCGCAAC-----GGTCTCCGGGA 1030
QY 47 ProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsnLys 66
DB 1031 GAGAGGGCCCCAACGGGCTCCGCGCTCCCTGGACGAGCGGCTCCAAAGCGCAGAG 1090
QY 67 GlyGlnLysGlyGluLysGlyProGlyProGlyProGlyProAlaGlyGluArgGlyPro 86
DB 1091 GAGAACGGGGCAGAGCTGGGAGCTGGTGAGCGCGCCCTCTGGAGAGCCAGCGCTC 1150
QY 87 IleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySer----- 104
DB 1151 CCTGGAGATGCTGGCATGCTGGGAGCGCGGTGAGCGGTGAGCGCGCTCAGCGGG 1210
QY 105 -----GlnGlyProLys 108
DB 1211 GCCCTGGGCCACAGGCCCTCCCGAGCCCTCGTGTCGAGGCTTCAGGGGCCAGAG 1270
QY 109 GlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyPro 128
DB 1271 GGCAGATGGAGACCCCGGCTTCCAGGCCCCCGAGGCTCCGAGGTGACGTGGCGCAG 1330
QY 129 ProGlyProGlyLysGlyGluLeuProGlyProGlnGlyProGlyPheGlnGly 148
DB 1331 CGGGTCCGGAGGTGCCAAGGCCCTTAAGGGAGACCAGGGTATTGCAGGTTCCGAC 1390
QY 149 LeuGlnGlyThrValGlyGlu----- 155
DB 1391 CTTCCTGGGATAAAGGAGAACTGGGTCCACGGCGCTGGTCCGACCACCAAGAGAG 1450
QY 155 ----- 155
DB 1451 GGCAGTCGAGGGAGGTGGGCCCCAAAGGCCACCCAGGGTCCCAACGGCACCAGCG 1510
QY 156 -----ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly 172
DB 1511 CAGGTTGTCGCCGGCCCCCGCTCTCTGGCCCTCGAGGGGCTCCCGGGTTCCTG 1570
QY 173 Met----- 173
DB 1571 ATCAGGGGAACCGGGAGTTCGGGGGAAGAGGCCAGCGGATCAGGGAGCTG 1630
QY 173 ----- 173
DB 1631 TGTGGGGATCATCAGCGAACAAATTCACAGTTAGCGCGCACCTTAAGAGCCTT 1690
QY 174 -----ProGlyProLysGlyProGlyProGlyProGlyProSer 186
DB 1691 GCACCGGGTCCATGTCGGCGCGCTCCAGCTGCCGCCCTGGCGGCCCGCCAGCCCA 1750
QY 187 GlyAlaVal-ValProLeuAlaLeuGlnAsnGluPro----- 198
DB 1751 GGCTCATTTGGTACCTGGCTCGAGGACCCCGGATACCGCGGTCCTCCAGGGAG 1810
QY 199 -----ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAs 216
DB 1811 CTGGGAGACCCCGGCCAGAG----- 1832
QY 216 pLysTyrTrpPheSerValGluGluLeuPheGluAspAlaLysLeuPheCys 236
DB 1833 -----GAAACAGGGTGACAGAGGAGACAAAG----- 1859
QY 236 uAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLys 256
DB 1860 -----CGCGGGCAGGAGCAG----- 1874

```

```

QY 256 sGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnG 276
DB 1875 -----GGCTGGACGGCCTGAAGGAGACCGAGGGCCCCAAGGA 1912
QY 276 uTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnPro 296
DB 1913 CCCCAGGC-GTGCCCGGCACCAAGCAGCGCAGGACGGTGTCTCCCGCGAGCCTGG 1971
QY 296 pAsnTrpGlyHisGlyHisGlyProGly----- 305
DB 1972 GCCTCCCGAGATCTCGGCTTCAGGTGCCATTGGGGCCCAAGGACACCGGGATCTG 2031
QY 306 -----GluAspCysAlaGlyLeuIleTrpAlaGly 315
DB 2032 CGACACCTCAGCCTGCCAAGGAGCGGTGTAGAGGG 2068

RESULT 8
US-09-073-663-11
; Sequence 11, Application US/09073663
; Patent No. 6127523
; GENERAL INFORMATION:
; APPLICANT: Brewton, Richard G.
; APPLICANT: Mayne, Richard
; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McGregor & Adler, LLP
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS 8.1
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,663
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benjamin Aaron Adler, Ph.D., J.D.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5913D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 777-2321
; TELEFAX: (713) 777-6908
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..2098
US-09-073-663-11

Alignment Scores:
Pred. No.: 8,17e-15 Length: 2543
Score: 419.00 Matches: 121
Percent Similarity: 35.59% Conservative: 26
Best Local Similarity: 29.30% Mismatches: 91
Query Match: 21.60% Indels: 176
DB: 3 Gaps: 9

US-09-763-712a-2_COPY_206_547 (1-342) x US-09-073-663-11 (1-2543)

QY 27 ValAspSerLysHisGlyGlnLeuLeuLysAsnPheThrIleLeuGlnGlyProGly 46
DB 986 CTCGATGGCCAGAGGAGAGCTGGTCGCAAC-----GGTCTCCGGGA 1030

```

Db 986 CTCGATGCCAAGAGAGAGAGCGTGTCCGAAC-----GGTCTCCGGGA 1030
 Qy 47 ProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsnLys 66
 Db 1031 GAGAGGGCCCAACAGCGCTCCGGGCTCCCTGACAGCGGGGTCCAAAGGCGAGAG 1090
 Qy 67 GlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGlyGluArgGlyPro 86
 Db 1091 GGAGAACGGGCGAGAGCTGGGAGCTGGTGAGGCGCGCCCTCTGTGAGAGCCAGCGCTC 1150
 Qy 87 lIeglyProAlaGlyProGlyGluArgGlyGlyGlySerLysGlySer----- 104
 Db 1151 CTTGGAGATGCTGGCATGCTCCGGAGCGGTGAGGCTGGCCACCGGGGTTCACGGGG 1210
 Qy 105 -----GlnGlyProLys 108
 Db 1211 GCCCTCGCCCAACAGGCGCTCCCGAGCGCTGTGTCCGAGGCTTCAGGCGCCAGAG 1270
 Qy 109 GlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyPro 128
 Db 1271 GGCAGCATGGGAGACCCCGGCTCCAGGCGCCAGGCGCTCCGAGGTGACGTGGCGAC 1330
 Qy 129 ProGlyProProGlyLysGluLysGlyLeuProGlyProGlnGlyProGlyPheGlnGly 148
 Db 1331 CGGGTCCGGGAGGTGCCGAAGCCCTAAGGAGACAGGATATTGACAGGTTCGCACGTT 1390
 Qy 149 LeuGlnGlyThrValGlyGlu----- 155
 Db 1391 CTTCTGGGATAAGGAGAACTGGGTCCAGCGCGCTGGTCCGAGCCCAAGAGGAGTCT 1450
 Qy 155 ----- 155
 Db 1451 GGCAGTCGAGGGAGCTGGGCCCCAAAGGACCCAGGCTCCCAAGCGCACCGCGTGT 1510
 Qy 156 -----ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly 172
 Db 1511 CAGGTGTCCCGGGCCCCCGGTCTCTGGGCTTCGAGGCGTCCCGGGTGTCTCTGGC 1570
 Qy 173 Met----- 173
 Db 1571 ATCAGGGGAAGCCGGAGTTCCGGGGAAGGAGCCAGGAGCGCATCAGGAGCTG 1630
 Qy 173 ----- 173
 Db 1631 TGTGGGGGATGATCAGCGAACAATTCACAGTTAGCGCGCACCTAAGGAAGCCTTTG 1690
 Qy 174 -----ProGlyProLysGlyProGlyProGlyProGlyProGlyProSer 186
 Db 1691 GCACCCGGGTCCATTGGTCGGCCCGGTCCAGCTGGCCCGCTGGGCGCCCGAGGACCCCA 1750
 Qy 187 GlyAlaVal-ValProLeuAlaLeuGlnAsnGluPro----- 198
 Db 1751 GGCTCCATTGTCACCTTGGCGCTCGAGGACCCCGGATACCGCGGTCCCACTGGGAG 1810
 Qy 199 -----ThrProAlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAs 216
 Db 1811 CTGGGAGACCCCGGCCAGAG----- 1832
 Qy 216 pLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysG 236
 Db 1833 -----GAAACCAAGGTGACAGAGGAGACAAAG----- 1859
 Qy 236 uAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLys 256
 Db 1860 -----CGCGGAGGAGCAG----- 1874
 Qy 256 sGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsn 276
 Db 1875 -----GGCTGGACGGGCTGAAGGAGACACGAGGCGCCCAAGGA 1912
 Qy 276 uTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyClnProAs 296
 Db 1913 CCCCAGGC-GTGCCCGGACACAGGAGCGGCGGACGAGCGGTGTCTCCCGGCGGCTGG 1971

Qy 296 pAsnTrpGlyHisGlyHisGlyProGly----- 305
 Db 1972 GCCTCCCGGAGATCTCGGCTTCAGGTCCTCCATTGGGCCCCAGGGACACCGGGGATCTG 2031
 Qy 306 -----GluAspCysAlaGlyLeuIleTyrAlaGly 315
 Db 2032 CGACACCTCAGCCTGCCAAGGAGCGGTGTAGGAGGG 2068
 RESULT 9
 US-08-642-255-48
 ; Sequence 48, Application US/08642255
 ; Patent No. 5773249
 ; GENERAL INFORMATION:
 ; APPLICANT: CAPPELLO, Joseph
 ; APPLICANT: FERRARI, Franco A.
 ; TITLE OF INVENTION: High Molecular Weight Collagen-Like
 ; TITLE OF INVENTION: Protein Polymers
 ; NUMBER OF SEQUENCES: 135
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/642,255
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: ROWLAND, Bertram I.
 ; REGISTRATION NUMBER: 20,015
 ; REFERENCE/DOCKET NUMBER: A55556-3/BI
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 494-8700
 ; TELEFAX: (415) 494-8771
 ; TELEX: 910 277299 FHT UR
 ; INFORMATION FOR SEQ ID NO: 48:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 432 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: other nucleic acid
 ; DESCRIPTION: /desc = "synthetic"
 ; US-08-642-255-48
 Alignment Scores:
 Pred. No.: 3.28e-15 Length: 432
 Score: 412.50 Matches: 80
 Percent Similarity: 55.70% Conservative: 3
 Best Local Similarity: 53.69% Mismatches: 57
 Query Match: 21.26% Indels: 9
 Db: 1 Gaps: 3
 US-09-763-712A-2_COPY_206_547 (1-342) x US-08-642-255-48 (1-432)
 Qy 43 GlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 62
 Db 10 GCCCAGCAGGTCGGAAGGCGCGCATGCCAGGCGCCGAAAGGTGCGCTGACCG 69
 Qy 63 ThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGly 82
 Db 70 GCTGTCCACCGGCTCTCCGGGACCTCGAGGCCCGCCAGGTGCGCTGACCGGCTGT 129
 Qy 83 GluArgGlyProLleGlyProAlaGlyProGlyProGlyGluArgGlyLysGlySerLys 102

Db 130 CCACGGGTCTCCGGACCTCCAGCGCCGCGAGGTGGCCCTGGACCGGCTGGTCCACCG 189
Qy 103 GlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyPro 122
Db 190 GGTCTCCGGACCT- - - - -GCAGCGCCGCGAGGTGGCCCTGGACCGGCTGGTCCA 240
Qy 123 SerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGlnGly 142
Db 241 CCGGGTGTCTCCGGACCTCCAGCGCCGCA- - - - -GGTGGCTGGACCGGTGTGT 291
Qy 143 ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArg 162
Db 292 CCACCG- - - - -GGTCTCCGGACCTGCAGCGCCGCGAGGTGGCCCTGGACCGGT 342
Qy 163 GlyLeuProGlyLeuProGlyValProGlyMetProGlyLysGlyProGlyProGlyPro 182
Db 343 GTCCACCGGGTCTCCGGACCTGCAGCGCCGCGAGGTGGCCCTGGACCGGTGGCCACGACCG 402
Qy 183 ProGlyProSerGlyAlaValValPro 191
Db 403 AAGGAGCTCAGGTCCGCGCAGGTCCG 429

RESULT 10

US-08-159-784-4
; Sequence 4, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3394
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-159-784-4
Alignment Scores:
Pred. No.: 3.75e-14
Score: 409.00
Percent Similarity: 48.98%
Best Local Similarity: 43.88%
Query Match: 21.08%
DB: 1
Length: 3394
Matches: 86
Conservative: 10
Mismatch: 42
Indels: 58
Gaps: 7

US-09-763-712a-2_COPY_206_547 (1-342) x US-08-159-784-4 (1-3394)
Qy 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySer- - - - - 56
Db 553 ATGCCCGCGCCCGAGGACCTCCAGGCGCCCGAGGCGCTCCAGGACTCTCTTTACGAC 612
Qy 57 - - - - -GlnGlyProProGlyProThrGlyAsnLysGly 67
Db 613 AGCAATGTGTTTGTCTAGTCCAGCGCCCGCGGCTCCAGGATTGCCAGGAATCAG- - - 669
Qy 68 GlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGlyProIle 87
Db 670 - - - - -GGCCCTCCAGGACCCCAAGGGCCCAAGAGAGAG 705
Qy 88 GlyProAlaGlyProGlyGlu- - - - - 95
Db 706 GGCCCGCCCGGACCAAGGAGTTCGGTTTGTACTTTTTCAGAGGAGGCTGAAATG 765
Qy 96 ArgGlyGlyLysGlySerLysGlnGlyProLysGlySerArgGlySer- - - - - 113
Db 766 AAGGGGAGAGGAGACCGAGGTGATGCAGGACAGAAAGGCGAAAGGGGGAGCGCGG 825
Qy 114 - - - - -ProGlyLysProGlyProGlnGlyProSerGly 124
Db 826 GCGCGCGGTTCCTCGGCTCCAGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCG 885
Qy 125 AspProGlyProGlyProGlyLysGlu- - - - -GlyLeuProGly 139
Db 886 TACCCTGGATTCCAGTCCCAAGGAGAGAGACATCCGGGGCCAGCGCCCGCCACTGGA 945
Qy 140 ProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValPro 159
Db 946 CCTCAGGACCGCCCGGCATC- - -GGCTACGAGGGCGCGAGGCGCTCCCGCGCCCA 1002
Qy 160 GlyProArgGlyLeuProGlyLeuProGly- - - - -ValProGly 172
Db 1003 GGCG 1062
Qy 173 MetProGlyProLysGlyProGlyProGlyProGlyProSerGlyAla 188
Db 1063 CCTCG 1110

RESULT 11

US-07-621-091G-1
; Sequence 1, Application US/07621091G
; Patent No. 5424408
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T., Morrison, Karen E., Hudson, Billy
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University, Office of Cooperative Research
; STREET: 246 Church Street
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800K storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS7.0
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/621,091G
; FILING DATE: 11/30/90
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA: No. 5424408 applicable
; ATTORNEY/AGENT INFORMATION:
; NAME: Barth, Richard S.
; REGISTRATION NUMBER: 28180
; REFERENCE/DOCKET NUMBER: 900983/RB

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 972-1400
TELEFAX: (212) 370-1622
TELEX: 236268
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1416 base pairs
TYPE: Nucleic acid
STRADEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Calf
STRAIN: Unknown
INDIVIDUAL ISOLATE: Unknown
DEVELOPMENTAL STAGE: Unknown
CELL TYPE: Whole kidney
IMMEDIATE SOURCE:
LIBRARY: Bovine lens cDNA
CLONE: KMC15
POSITION IN GENOME: No. 5424408 known
CHROMOSOME/SEGMENT: No. 5424408 known
PUBLICATION INFORMATION: No. 5424408e
US-07-621-091G-1

Alignment Scores:
Pred. No.: 3.6e-14 Length: 1416
Score: 402.50 Matches: 103
Percent Similarity: 46.62% Conservative: 21
Best Local Similarity: 38.72% Mismatches: 103
Query Match: 20.75% Indels: 40
DB: 1 Gaps: 8

US-09-763-712A-2_COPY_206_547 (1-342) x US-07-621-091G-1 (1-1416)

```
QY 42 GlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGly 61
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 112 AAAGGAGATCGAGGTCCACCTGGCTCCAGAGAACCCAGGATGCTGCTCTCTGA 171

QY 62 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 76
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 CCTCAGGAGGAGTCTAGAACGCATATAAAGGAGACAGGGTGTGATGGAGAGCCTGCG 231

QY 77 ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArg 96
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 CAAAGAGGTCCACCTGGAGCATATAGGAGATGCTGCTGCTGCTGCTGCTGCTG 291

QY 97 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 116
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 GGTGTCCCGGTGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 351

QY 117 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyLysGluGly 136
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 352 AAAGGAGAGAGGGTAATTCAGGATTTCCAGGACCATCCAGGAGGAGGAGGAGG 411

QY 137 -----LeuPro 138
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 412 CCAAAAGGACCATCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 471

QY 139 GlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 158
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 472 GGAAGCCAGGCCACCTGCTGAGTGGAGAACCCAGGATGCAAGAGAACCCGGGCC 531

QY 159 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 178
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 532 CCAGGACCAACAGGAGATCCAGGACCTGTGGGCCAAAAGGTAAACACGAGGAGGATGT 591

QY 179 ProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 198
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 592 CCACCAAGGAACCTGGAGAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 651
```

```
QY 199 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPro 214
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 652 GGACCA---CCTGGATCCGATGCGCTGCCAGGCTTCAAGGGGAACCC-TGGAGACACTGG 707

QY 215 ThrAspLysCysTrpTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 234
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 708 ACC---ACCTGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 758

QY 235 CysGluAspLysSerSerHisValPheIleAsnThrArgGluGlnGlnTrpIle 254
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 759 ---CCAGACCAACAGCAATCCCTCTCTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 815

QY 255 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 272
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 816 TTCTCTCTCTCTTTGTACAAAGAAATGAACAAGCCCATGGACAGGA-----850

QY 273 ArgGluAsnGluTrpLys 278
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 861 CCTGGGAACACTTGGCAG 878
```

RESULT 12

US-08-399-889-1
Sequence 1, Application US/08399889B
Patent No. 5973120
GENERAL INFORMATION:
APPLICANT: Reiders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263A
CURRENT APPLICATION NUMBER: US/08/399,889B
CURRENT FILING DATE: 1995-03-07
EARLIER APPLICATION NUMBER: 07/621091
EARLIER FILING DATE: 1990-11-30
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1416
TYPE: DNA
ORGANISM: Calf
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1416)
US-08-399-889-1

Alignment Scores:
Pred. No.: 3.6e-14 Length: 1416
Score: 402.50 Matches: 103
Percent Similarity: 46.62% Conservative: 21
Best Local Similarity: 38.72% Mismatches: 103
Query Match: 20.75% Indels: 40
DB: 1 Gaps: 8

US-09-763-712A-2_COPY_206_547 (1-342) x US-08-399-889-1 (1-1416)

```
QY 42 GlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGly 61
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 112 AAAGGAGATCGAGGTCCACCTGGCTCCAGAGAACCCAGGATGCTGCTCTCTGGA 171

QY 62 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 76
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 CCTCAGGAGGAGTCTAGAACGCATATAAAGGAGACAGGGTGTGATGGAGAGCCTGCG 231

QY 77 ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArg 96
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 CAAAGAGGTCCACCTGGAGCATATAGGAGATGCTGCTGCTGCTGCTGCTGCTG 291

QY 97 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGly 116
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 GGTGTCCCGGTGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 351

QY 117 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyLysGluGly 136
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 352 AAAGGAAGAAGGTAATTCAGGATTTCAGGACCACCTGGACCTCCAGGGCAAAAGTGA 411
QY 137 -----LeuPro 138
Db 412 CCAAAAGGACCACCTGGAGTACGTGGAGACCTGGGCACAGTGAAGATCATCTCCCTTCCA 471
QY 139 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 158
Db 472 GGAAGGCCAGGCCACCTGGTTCAGCTGGAGAACACAGGATGCAAGGAGAACCCGGGCC 531
QY 159 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProGlyGly 178
Db 532 CCAGGACCCAGGAGATCAGGACCTGTGGGCCCAAAAGGTAACACAGGGAGGATGGT 591
QY 179 ProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 198
Db 592 CCACCAAGGACTCTCGACCACTGGAGAAAAGGCAACAAAGTTGTAAAGGAGAGCAA 651
QY 199 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpIleGlnTrpIle 214
Db 652 GGACCA---CCTGGATCCGATGGCTGCCAGGCTTGAAGGGGAAACC-TGGAGACACTGG 707
QY 215 ThrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 234
Db 708 ACC---ACCTGCAGAGGGCAGTGTATGAGGGCTTTGTCTTTACCCGGCAGACAG----- 758
QY 235 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 254
Db 759 ---CCAGACCCAGCAATTCCTCTGTCCAGNAGGAGCAGACCCCTCTATAGTGGTT 815
QY 255 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 272
Db 816 TTCTCTCTCTTTGTACAAAGGAAATGAACAGCCCATGGACAGGA-----860
QY 273 ArgGluAsnGluTrpLys 278
Db 861 CTGGGAACACTTGGCAG 878
```

RESULT 13

US-09-167-364-1

; Sequence 1, Application US/09167364

; Patent No. 6007980

; GENERAL INFORMATION:

; APPLICANT: Reeders, Stephen T

; APPLICANT: Morrison, Karen E

; APPLICANT: Hudson, Billy G

; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

; FILE REFERENCE: 951263B

; CURRENT APPLICATION NUMBER: US/09/167,364

; CURRENT FILING DATE: 1998-10-07

; EARLIER APPLICATION NUMBER: 08/399889

; EARLIER FILING DATE: 1995-03-07

; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1416

; TYPE: DNA

; ORGANISM: Calf

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1416)

US-09-167-364-1

Alignment Scores:

| | | | |
|------------------------|---------|---------------|------|
| Pred. No.: | 3-6e-14 | Length: | 1416 |
| Score: | 402.50 | Matches: | 103 |
| Percent Similarity: | 46.62% | Conservative: | 21 |
| Best Local Similarity: | 38.72% | Mismatches: | 103 |
| Query Match: | 20.75% | Indels: | 40 |
| DB: | 3 | Gaps: | 8 |

US-09-763-712A-2_COPY_206_547 (1-342) x US-09-167-364-1 (1-1416)

```
QY 42 GlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGly 61
Db 112 AAAGGAGATCGAGGTCCACCTGGCTCCAGAGGAAACAGGATGCTGCTGCTCCTCTGGA 171
QY 62 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 76
Db 172 CTCCAGGAGTCTCTGTAGAAAGGAGCAAGGGTTGATGGAGAGCCTGGC 231
QY 77 ProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArg 96
Db 232 CAAAGAGTCCACCTGGAGCTATAGGAGCATGGGTCCACAGTCTATCGGGAGCACA 291
QY 97 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 116
Db 292 GGTGTCCCGGTCAGCCAGGCGGCAGAGGTGATCTCTGATTTCCAGGACATG 351
QY 117 ProGlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGluGly 136
Db 352 AAAGGGAAGAAGGTAATTCAGGATTTCCAGGACCACTGGACCTCCAGGCAAAAGTGA 411
QY 137 -----LeuPro 138
Db 412 CCAAAAGGACCACCTGGAGTACGTGGAGACCTGGGCACAGTGAAGATCATCTCCCTTCCA 471
QY 139 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 158
Db 472 GGAAGGCCAGGCCACCTGGTTCAGCTGGAGAACACAGGATGCAAGGAGAACCCGGGCC 531
QY 159 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 178
Db 532 CCAGGACCCAGGAGATCAGGACCTGTGGGCCCAAAAGGTAACACAGGGAGGATGGT 591
QY 179 ProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 198
Db 592 CCACCAAGGACTCTCGACCACTGGAGAAAAGGCAACAAAGTTGTAAAGGAGAGCAA 651
QY 199 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 214
Db 652 GGACCA---CCTGGATCCGATGGCTGCCAGGCTTGAAGGGGAAACC-TGGAGACACTGG 707
QY 215 ThrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 234
Db 708 ACC---ACCTGCAGAGGGCAGTGTATGAGGGCTTTGTCTTTACCCGGCAGACAG----- 758
QY 235 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 254
Db 759 ---CCAGACCCAGCAATTCCTCTGTCCAGAAAGGAGCAGACCCCTCTATAGTGGTT 815
QY 255 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 272
Db 816 TTCTCTCTCTTTGTACAAAGGAAATGAACAGCCCATGGACAGGA-----860
QY 273 ArgGluAsnGluTrpLys 278
Db 861 CTGGGAACACTTGGCAG 878
```

RESULT 14

US-09-439-897-1

; Sequence 1, Application US/09439897

; Patent No. 6277558

; GENERAL INFORMATION:

; APPLICANT: Hudson, Billy G

; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

; FILE REFERENCE: 95-1263-C

; CURRENT APPLICATION NUMBER: US/09/439,897

; CURRENT FILING DATE: 1999-11-12

; NUMBER OF SEQ ID NOS: 65

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1416

; TYPE: DNA

; ORGANISM: Bos taurus

; FEATURE:

```
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
US-09-439-897-1

Alignment Scores:
Pred. No.: 3,6e-14 Length: 1416
Score: 402.50 Matches: 103
Percent Similarity: 46.62% Conservative: 21
Best Local Similarity: 38.72% Mismatches: 103
Query Match: 20.75% Indels: 40
DB: 4 Gaps: 8

US-09-763-712A-2_COPY_206_547 (1-342) x US-09-439-897-1 (1-1416)
QY 42 GlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGly 61
DB 112 AAAGGAGATCGAGTCCACCTGGCTCCAGAGGAACCCAGGATCGCTGCTCTCGA 171
QY 62 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 76
DB 172 CTTCCAGGGAGTCTGTAGAGGCATAAAGAGACAAGGGTTGATGGAGAGCTGCG 231
QY 77 ProProGlyProAlaGlyGluArgGlyProLleGlyProAlaGlyProProGlyGluArg 96
DB 232 CAAGAGGTCCACCTGGAGCTATAGAGACATGGGGTCCACAGGTCTATCCGGGACACCA 291
QY 97 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 116
DB 292 GGTGTCCCGGTACCCAGGGCCAGAGGTGATCTGTGATCTATGGATTTCCAGGCATG 351
QY 117 ProGlyProGlnGlyProSerGlyAspProGlyProProGlyProGlyLysGluGly 136
DB 352 AAAGGGAAGAGGTAAATTCAGGATTTCCAGGACCACTGGACCTCCAGGGCAAAAGTGA 411
QY 137 -----LeuPro 138
DB 412 CCAAAAGGACCACTGGAGTACGTGGAGAGCTGGCAGACAGTGAAGATCATCTCCCTTCCA 471
QY 139 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 158
DB 472 GGAAGCCAGGCCCCACCTGGTTCAGCTGGAGAACACAGGAGTCAAGGAGAACCCGGGCC 531
QY 159 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 178
DB 532 CCAGGACCAACAGGAGATCCAGGACCTCTGGGCCCAAAAGGTAAACACAGGGAGGATGT 591
QY 179 ProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 198
DB 592 CCACCAAGAACTCTCGGACCACTGGAGAAAAAGCAAAAGGTGTGAAGGAGAGCA 651
QY 199 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTriLysAsnPhe 214
DB 652 GGACCA---CCTGGATCCGATGGCTCGCAGCTTGAAGGGGAAACC-TGGAGACACTGG 707
QY 215 ThrAspLysCysTyrrPheSerValGlyLysGluIlePheGluAspAlaLysLeuPhe 234
DB 708 ACC---ACCTGCAGCAGGGCGAGTATGAGGGCTTTGTCTTTACCGGSCACAG----- 758
QY 235 CysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIle 254
DB 759 ---CCAGACCAACAGAAATTCCTCTCTCCAGAGGGGACAGCGCGCTCTATAGTGGGTT 815
QY 255 LysLysGlnMetValGlyArgGlu-----SerHisTriLleGlyLeuThrAspSerGlu 272
DB 816 TTCTCTTCTTTGTACAGGAATTAACACGCCCACTGACACAGGA----- 860
QY 273 ArgGluAsnGluTrpLys 278
DB 861 CTGGGAACACTTGGCAG 878
```

RESULT 15

US-09-484-970B-4

; Sequence 4, Application US/09484970B

```
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 4359
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 331365 (1553795CBL)
US-09-484-970B-4
```

```
Alignment Scores:
Pred. No.: 1,38e-13 Length: 4359
Score: 400.50 Matches: 90
Percent Similarity: 54.89% Conservative: 11
Best Local Similarity: 48.91% Mismatches: 64
Query Match: 20.64% Indels: 19
DB: 4 Gaps: 7

US-09-763-712A-2_COPY_206_547 (1-342) x US-09-484-970B-4 (1-4359)
```

```
QY 41 LeuGlnGlyProProGlyProArgGlyProArgGlyAspArg----- 54
DB 448 CTGCCAGGACCCAGGCCACAGGTGCCAAAGGCTCCGAGGAATCCAGGCTTCGCA 507
QY 55 GlySerGlnGlyProProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGlu 74
DB 508 GGAGTGTATGGAGGACCAAGGCTTCCAGGAGTCCAGGAGACGACGATCGTGAAGGTT 567
QY 75 ProGlyProPro-----GlyProAlaGlyGluArgGlyProLleGlyProAlaGly 91
DB 568 CCAGGACCCAGGTTTCATAGACCCCGAGGATCCAAAGGTGCATGGGCTCCCTGGC 627
QY 92 ProProGlyGluArgGlyLysGlySerLysGlySerGlnGlyProLysGlySerArg 111
DB 628 CCAGATGGATCCAGGTCCCATCGGCTCCAGGCGCCAGATGGGCGCCCTGGGAAAG 687
QY 112 GlySerProGlyLysPro---GlyProGln---GlyProSerGlyAspProGlyProPro 129
DB 688 GGCTCTCCCTGGAGAGTCTCTGGAGCTCAGCCCGGCCACGGGAGATGCTGGTGGCT 747
QY 130 GlyProProGlyLysGluGlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeu 149
DB 748 GGACAGCTTGGCTTAAAGGCTTCCCGAGACAGAGGCCCTCGCTGGATTCAGAGAAC 807
QY 150 GlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGly 169
DB 808 CAAGGATG-----CCTGGATGTCAGGCTCAAGGCGCCAGGCCAGGCTCCACAGA 858
QY 170 ValProGlyMetProGlyProLysGlyProProGlyPro-----ProGlyProSer 186
DB 859 CCTTCGGCCAGCAGGCTGTATGGGCTCCAGGACTGCATGGATTCAGGAGTCCT 918
QY 187 GlyAlaValValProLeuAlaLeuGlnAsnGluPro-----ThrProAlaProGluAsp 204
DB 919 GGCCCAAGAGGGGCTTGGGGCTGCCAGGAATCCCGAGCGCTGAGAGTCTGCTGTGAT 978
QY 205 AsnGlyCysPro 208
DB 979 AGAGGGGACCT 990
```

Search completed: March 21, 2003, 12:34:27

Job time : 46.3343 secs

| Result No. | Score | % Match | Length | ID | Description |
|------------|-------|---------|--------|----|--------------------|
| 1 | 1934 | 99.7 | 2930 | 10 | US-09-745-763-198 |
| 2 | 474.5 | 24.5 | 1410 | 10 | US-09-954-456-2274 |
| 3 | 455 | 23.5 | 1133 | 9 | US-09-924-340-57 |
| 4 | 455 | 23.5 | 1133 | 9 | US-09-992-600A-57 |

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/745,763
; FILING DATE: 18-Jun-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 198:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2930 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-09-745-763-198

Alignment Scores:
Pred. No.:          Length:      2,49e-95      2930
Score:              1934.00      Matches:      341
Percent Similarity: 99.71%      Conservative: 0
Best Local Similarity: 99.71%      Mismatches: 1
Query Match:        99.69%      Indels:      0
DB:                  10         Gaps:        0

US-09-763-712A-2_COPY_206_547 (1-342) x US-09-745-763-198 (1-2930)
QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
Db 1266 ATGCAACAAGATTGATGAGGTGCGAGGTTAGACACTGGAAGTAGCCCAACTATCAGTGATT 1325
QY 21 MetGluGluMetLysLeuValAspSerLysHISGlyGlnLeuIleLysAsnPheThrIle 40
Db 1326 ATGGAAGAATGAGCTAGTAGCTCAAGTAGCTGCTCATCAAGAATTTTACAATA 1385
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
Db 1386 CTACAAGGTCCACCGGGCCCCAGGGGTCCCAAGAGGTGACAGAGGATCCCAGGGACCCCT 1445
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
Db 1446 GCGCCCAACTGGCAACAAGGACAGAAAGAGAGAGAGGGGGGCGCTGGACACCTGGCCCT 1505
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGlyClyuArgGlyLysGly 100
Db 1506 GCGGGTGAGAGAGGCCCAATTGACAGCTGGTCCCGGAGAGCGTGGCGGCAAGGA 1565
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
Db 1566 TCTAAAGGCTCCAGGGCCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGGCCCTCAG 1625
QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyGlyGluGlyLeuProGlyPro 140
Db 1626 GCGCCCACTGGGACCCAGCGCCCCCGGCCCAAGGAGAGAGGAGGAGGAGGAGGAGGAGG 1685
QY 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyClyuProGlyValProGly 160
Db 1686 CAGGGCCCCCTGGCTTCAGGGACTTCAGGGACCGTTGGGAGCGCTGGGGTGCCCTGGA 1745
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
Db 1746 CCTCGGGAGTCCAGGCTTGCTGGGTACCAGGCATCCAGGCCCCCAAGGGCCCCCCC 1805
QY 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
Db 1806 GGCCCTCTCTGGCCCATCAGGAGCGGTGGTGCCTCGCCCTGCAGAATCAGCAACCCCG 1865
QY 201 AlaProGluAspAsnGlyCysProProHISLysThrLysAsnPheThrAspLysCysThr 220
Db 1866 GCACCGGAGGACAAATAGCTGCCCGCTCCTACTGGAAGAACTTCACAGACAAATGCTACT 1925
```

```

QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSer 240
Db 1926 TTTTCAGTTGAGAAACAAATTTTGGAGTAGCAAGCTTTCTGTGAAGACAAGTCTTCA 1985
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
Db 1986 CATCTTTGTTTATATAACACTAGAGAGAGACACAGCAATGGATAAAAAACAGATGGTAGG 2045
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
Db 2046 AGAGAGAGCCACTGGATCGCCCTCACAGACTCAGAGCGGTGAAAAATGAATGAAGTGCGTG 2105
QY 281 AspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
Db 2106 GATGGGACATCTCCACACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGTCAT 2165
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
Db 2166 GCCATGGGCCAGGAGAAAGACTGTGCTGGGTTGATTTATGCTGGGAGTGAACGATTTC 2225
QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
Db 2226 CAATGTGAAGACGTCAATTAATTCATTTCGAAAAAGACAGAGGAGACAGTACTGTCTCT 2285
QY 341 AlaLeu 342
Db 2286 GCATTA 2291

RESULT 2
US-09-954-456-2274
; Sequence 2274, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2274
; LENGTH: 1410
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-2274

Alignment Scores:
Pred. No.:          Length:      6.9e-18      1410
Score:              474.50      Matches:      111
Percent Similarity: 49.67%      Conservative: 41
Best Local Similarity: 36.27%
```

| | | | | |
|---|--------|---|---------|----|
| Query Match: | 24.46% | | Indels: | 23 |
| DB: | 10 | | Gaps: | 6 |
| US-09-763-712A-2_COPY_206_547 (1-342) x US-09-954-456-2274 (1-1410) | | | | |
| Qy | 41 | LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro | 60 | |
| Db | 400 | ATGCCTGGACAGCTGCCCGATTGGCCCAAGGGACAATGGCTCTGTGGGAACCT | 459 | |
| Qy | 61 | GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyPro | 80 | |
| Db | 460 | GGACCAAAGGAGACACTGGCCAAAGTGGACCTCCAGGACCTCCCGGTGTCCCTGGTCCA | 519 | |
| Qy | 81 | AlaGlyGluArgGlyProIleLeuGlyProAlaGlyProProGlyGluArgGlyGlyLysGly | 100 | |
| Db | 520 | GCTGGAGAAGAGTCCCCTGGGAAGCAGGGACAATAGGACCTCAGGGCAAGCCAGGC | 579 | |
| Qy | 101 | SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln | 120 | |
| Db | 580 | CCAAAAGGAGAAGCTGGCCCCAAGGAGAAGTAGTGCCCCAGGCATCCAGGGCTCGCCA | 639 | |
| Qy | 121 | GlyProSerGlyAspProGlyProProGlyPro-----ProGlyLysGluGlyLeu | 137 | |
| Db | 640 | GGGGCAAGAGCCCTCCAGCGCCTCAAGGAGCAGGAGTGTCCCTGGTGTAGCGTGGAGTC | 699 | |
| Qy | 138 | ProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGly | 157 | |
| Db | 700 | CCTGGAAACCGCAGGGCACAGGCTGCTGGAGCCATGGGTCCCAGGGAAGTCCAGGT | 759 | |
| Qy | 158 | ValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys | 177 | |
| Db | 760 | GCACAGGGACCCCGGATTGAAGGGGACAAAGGCATCTCTGGAGACAAAGGACCAAG | 819 | |
| Qy | 178 | GlyProProGlyProProGlyProSerGlyAlaValValProLeu--AlaLeuGlnAsn | 196 | |
| Db | 820 | GGAGAAGTGGGCTCCAGATGTCCTCTCGAGGCAGCAGGTGTAGCGCTTACAGGGA | 879 | |
| Qy | 197 | GluProThrProAlaProGluAspAsnGlyCys----- | 207 | |
| Db | 880 | CAAGT-ACAGCACCTCCA-----GGCTGCTTTCTCTCAGTATAAGAAAGTTGAGCT | 929 | |
| Qy | 208 | ProProHisTrp-LysAsnPheThrAspLysCysTyrtYrPheSerValGluLysGluIl | 227 | |
| Db | 930 | CTTCCCAATGGCCAAAGTGTGGGAGAGAATTTCAAGACAGCAGCGCTTTGTAAACC | 989 | |
| Qy | 227 | ePheGluAspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIlLeAsnTh | 247 | |
| Db | 990 | ATTTACGGAGGCACAGCTGCTGTGCACACAGCGCTGGTGACAGTGTGGCTCTCCACGCTC | 1049 | |
| Qy | 247 | rArGluGluGlnGlnTrpIleLysLysGlnMetValGlyArg---GluSerHisTrpIl | 266 | |
| Db | 1050 | TGCCGCTGGAATPGCCCGCTTCCAACAGCTGGCTGTAGCTAAGAACAGCGGTGCTTTCCT | 1109 | |
| Qy | 266 | eGlyLeuThrAspSerGluArgGluAsnGlnTrpLysTrpLeuAspGlyThrSerProAs | 286 | |
| Db | 1110 | GAGCATCACTGATTCCAACAGACAGGCGNAGTTCACCTACCCACAGGAGAGTCCCTGGT | 1169 | |
| Qy | 286 | pTYrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGl | 306 | |
| Db | 1170 | CTATTCCAACTGGGCCCGCCAGGGGAGCCCAACGAT-----GATGGCGGGTCAGA | 1217 | |
| Qy | 306 | uASPcysAlaGlyLeuIleTyraLalGlyGlnTrpAsnAspPheGlnCysGluAspValas | 326 | |
| Db | 1218 | GGACTGTGTGGAGATCTTCCACCAATGGCAAGTGGAAATGACAGGCGCTGTGGAGAAAAAGCG | 1277 | |
| Qy | 326 | nAsnPheIleCysGlu | 331 | |
| Db | 1278 | TCTGTGGTCTGGAG | 1293 | |

RESULT 3

RESULT 3
US-09-924-340-57

03-09-924-340-37
: Sequence 57. Application IIS/09924340

; sequence 37, Application OS/099
; Publication No. US20030027248A1

```

; GENERAL INFORMATION:
; APPLICANT: Bejanin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

```

| | | | |
|----|------|--|------|
| Qy | 41 | LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro | 60 |
| Db | 400 | ATGCCTGGACAAGCTGGCCCAATGGGGCCCAAGAGGACAATGGCTCTGTGTGAGAACCT | 459 |
| Qy | 61 | GlyProThrGlyAsnLysGlnLysGlyGluProGlyProProGlyPro | 80 |
| Db | 460 | GGACCAAGGGAGACACTGGCCCAAGTGGACCTCCAGGACCTCCCGGTGTCTGGTCCA | 519 |
| Qy | 81 | AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly | 100 |
| Db | 520 | GCTGGAAAGAAAGTCCCTCGGGGAAGCAGGGGAACAATAGGACCTCAGGGCAAGCCAGGC | 579 |
| Qy | 101 | SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln | 120 |
| Db | 580 | CCAAAGAGAGAGCTGGGCCCAAGGAGAAGTAGTGGCCCGCAGCATGACAGGCTCGGCA | 639 |
| Qy | 121 | GlyProSerGlyAspProGlyProProGlyPro-----ProGlyLysGluGlyLeu | 137 |
| Db | 640 | GGGCAAGAGGCTCGCAGGCGCTTAAGGGAGAGCGAGGTCTCCTGTGTGAGCGTGGAGTC | 699 |
| Qy | 138 | ProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGly | 157 |
| Db | 700 | CCTGGAAACAGCGGGCAGCAGGGTCTGCTGGAGCCATGGGTCCCAAGGAAGTCCAGGT | 759 |
| Qy | 158 | ValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys | 177 |
| Db | 760 | GCCAGGGACCCCGGATGAAGGGGAGCAAAAGGCATTCTGGAGACAAGAGGACAAAG | 819 |
| Qy | 178 | GlyProProGlyProProGlyProSerGlyAlaValValProLeu-----AlaLeuGlnAsn | 196 |
| Db | 820 | GGAGAAAGTGGCTTCACAGATGTTGCTCTCTGAGGAGCAGGTTGAGGCTTACAGGGA | 879 |
| Qy | 197 | GluProThrProAlaProGluAspAsnGlyCys----- | 207 |
| Db | 880 | CAAGT-ACAGCACCTCCA-----GGCTGCTTCTCTCAGTATAAGAAAGTTGAGCT | 939 |
| Qy | 208 | ProProHisTrp-LysAsnPheThrAspLysCysTyTyTyPheSerValGluLysGluIl | 227 |
| Db | 930 | CTTCCCAAAATGGCCAAAGTGTGGGGAGAAGATTTTCAAGACAGCAGCGCTTGTATAAAC | 989 |
| Qy | 227 | ePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnTh | 247 |
| Db | 990 | ATTTACGGAGCAGCTGCTGTGCACACAGGCTGTGGAGACTGTGGCTCTCCAGCTC | 1040 |
| Qy | 247 | rArgGluGlnGlnTrpIleLysLysGlnMetValGlyArg-----GluSerHisTrpIl | 266 |
| Db | 1050 | TGCCGCTGAGAAATGCCCGCTTGAACAGCTGGTCGTAGCTAAGAACGAGCGCTGCTTTCCT | 1100 |
| Qy | 266 | eGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAs | 286 |
| Db | 1110 | GAGCATGACTGATTCCAAAGACAGAGGGCAAGTTCACCTACCCCAAGAGAGTCCCTGGT | 1160 |
| Qy | 286 | pTyLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyG | 306 |
| Db | 1170 | CTATTCCAACTGGGCCCGGAGGCCCAAGAT-----GATGGCGGTTCAGA | 1210 |
| Qy | 306 | uAspCysAlaGlyLeuIleTyAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAs | 326 |
| Db | 1218 | GGACTGTGTGGAGATTTCCAAATGGCAAGTGAATGACAGGCTGTGTGGAGAAAAGCG | 1278 |

RESULT 3

RESULT 3
US-09-924-340-57

03-09-924-340-37 : Sequence 57. Application IIS/09924340

; sequence 37, Application OS/099
; Publication No. US20030027248A1

```

: GENERAL INFORMATION:
: APPLICANT: Bejanin, Stephane
: APPLICANT: Tanaka, Hiroaki
: TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES T
: FILE REFERENCE: 91.US2.REG
: CURRENT APPLICATION NUMBER: US/09/924,340
: CURRENT FILING DATE: 2001-08-06
: PRIOR APPLICATION NUMBER: US 60/305,456
: PRIOR FILING DATE: 2001-07-13
: PRIOR APPLICATION NUMBER: US 60/302,277
: PRIOR FILING DATE: 2001-06-29
: PRIOR APPLICATION NUMBER: US 60/298,698
: PRIOR FILING DATE: 2001-06-15
: PRIOR APPLICATION NUMBER: US 60/293,574
: PRIOR FILING DATE: 2001-05-25
: NUMBER OF SEQ ID NOS: 112
: SOFTWARE: Jpatent
: SEQ ID NO 57
: LENGTH: 1133
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: 5'UTR
: LOCATION: 1..475
: NAME/KEY: CDS
: LOCATION: 476..964
: NAME/KEY: 3'UTR
: LOCATION: 965..1133
: NAME/KEY: polyA_signal
: LOCATION: 1101..1106
: NAME/KEY: polyA_site
: LOCATION: 1118..1133
: US-09-924-340-57

Alignment Scores:
Pred. No.: 6,26e-17 Length: 1133
Score: 455.00 Matches: 98
Percent Similarity: 47.46% Conservative: 14
Best Local Similarity: 41.53% Mismatches: 50
Query Watch: 23.45% Indels: 74
DB: 9 Gaps: 7

```

US-09-763-712A-2 COPY 206 547 (1-342) X US-09-924-340-57 (1-1133)

| | | | |
|----|-----|--|-----|
| Qy | 41 | LeuGlnGlyProProGlyProAArgGlyProAArgGlyAspArgGlySerGlnGlyProPro | 60 |
| | | | |
| Db | 140 | CTCCAGGCCCCCAGCCAGCGGGTCCAGAGGAGCGAGGACCCCAAGGTAACCTCC | 199 |
| | | | |
| Qy | 61 | GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyPro | 80 |
| | | | |
| Db | 200 | GGTGAGAAGGCGACAGGATTCAAGGCCAGCCAGGCTTTCGGGCCCCACCGGGTCCC | 259 |
| | | | |
| Qy | 81 | AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGly | 100 |
| | | | |
| Db | 260 | CCTGGATTCCAGGCAAGTTGGATCATCCTGGGCCACCTGGCCCTCAAGCAGAGAAGGCC | 319 |
| | | | |
| Qy | 101 | SerLysSerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyPro--- | 119 |
| | | | |
| Db | 320 | AGCGAAGGGATTGAGGCCCATCAGGCCCTGCCTGGCTCCCTGGGCCACCGGACCTCCT | 379 |
| | | | |
| Qy | 120 | -----GlnGlyProSer----- | 123 |
| | | | |
| Db | 380 | GGGATTTCAGGGCCCCCGCTCGGATGGTGGTGGGAAGGATGGCAAGCCTGGCTTG | 439 |
| | | | |
| Qy | 124 | ---GlyAspProGlyProProGlyProPro--- | 132 |
| | | | |
| Db | 440 | AGGGGGACCTTGCTCTGCTGGCCCCCTGGACTATGGGACCCACCGGGCTTTAAGGGG | 499 |
| | | | |
| Qy | 133 | -----GlyLysGluGlyLeuProGlyProGln----- | 144 |
| | | | |
| Db | 500 | AAACAGACGATCCTGGCTCCCCAGGACCTAAGGGTGACTGTGGCAACACGAGCTCCTCT | 559 |
| | | | |
| Qy | 145 | GlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArqGlyLeu | 164 |

| | | | | | |
|--|----------|--|------|--|--|
| Alignment Scores: | | | | | |
| Pred. No.: | 3,35e-16 | Length: | 6158 | | |
| Score: | 453.00 | Matches: | 130 | | |
| Percent Similarity: | 39.81% | Conservative: | 38 | | |
| Best Local Similarity: | 30.81% | Mismatches: | 76 | | |
| Query Match: | 23.35% | Indels: | 180 | | |
| DB: | 10 | Gaps: | 18 | | |
| US-09-763-712a-2_copy_206_547 (1-342) x US-09-919-497-6 (1-6158) | | | | | |
| QY | 43 | GlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 62 | | | |
| DB | 4032 | GGACCTCCAGGTGCCAAGGGCGCCAGGTGATGGCTTAAGGGTAACCGGGTCT 4091 | | | |
| QY | 63 | Thr----- | 63 | | |
| DB | 4092 | GTGGTTTCTGGAGATCCTGGTCTCTGGGAACTTGGCCCTGCAGGTCAAGTGT 4151 | | | |
| QY | 64 | -----GlyAsnLysGlnLysGlyGluLysGlyGluProGlyProGlyProAla 81 | | | |
| DB | 4152 | GTGGTGGTGACAAAGGTGAAGATGGATCCTGGTCAACCGGGTCTCTGGCCATCT 4211 | | | |
| QY | 82 | GlyGluArgGlyProIleGlyPro-----AlaGlyProProGlyGluArgGly 98 | | | |
| DB | 4212 | GGTGGGCTGGGCCACCGAGGTCTCTGGAAACGAGGTCCTCTGGAGCTGCAGGTGCA 4271 | | | |
| QY | 99 | LysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysPro--- 117 | | | |
| DB | 4272 | GAGGGAGACAGGTGAAGAGGTGCTAAGGGGAAGCAGGTGCAGAAAGTCTCTCTGGA 4331 | | | |
| QY | 118 | -----GlyProGlnGlyProSerGlyAspProGlyPro----- 128 | | | |
| DB | 4332 | AAAACGGCCAGTCTGGTCTCAGGACCTCGAGGAAGCCTGGTCCAGAAGTCTTCGG 4391 | | | |
| QY | 129 | -----ProGlyProProGlyGluGlyLeu----- 137 | | | |
| DB | 4392 | GGATCCCTGGTCTGGGGAACAAAGTCTCCCTGGAGCTGCAGGCCAAGATGAGCA 4451 | | | |
| QY | 138 | ProGlyProGlnGlyProGlyPhe----- 146 | | | |
| DB | 4452 | CCTGGTCTATGGGACCTCTGGCTACCTGGTCTCAAGGTGACCTGGCTCCAAGGT 4511 | | | |
| QY | 146 | ----- 146 | | | |
| DB | 4512 | GAAAGGGACATCTGTTTAAATTGGCCTGATTGGTCTCCAGGAGCAACAGGGGAAAA 4571 | | | |
| QY | 147 | -----GlnGlyLeuGlnGlyThrValGlyGluPro-----GlyVal 158 | | | |
| DB | 4572 | GTGACCGAGGGCTCCCTGGAATCAAGGATCTCCAGGAGCAAAAGGGATGGGGAATT 4631 | | | |
| QY | 159 | ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys--- 177 | | | |
| DB | 4632 | CCTGGTCTCTGGTCCCTTAGGTCCACCTGGTCTCCAGGTACCAAGGTCTCAAGGC 4691 | | | |
| QY | 177 | ----- 177 | | | |
| DB | 4692 | CCAAAGGGTAACAAAGGCTCTACTGGACCCGCTGGCCAGAAAGGTGACAGTGTCTTCCA 4751 | | | |
| QY | 178 | GlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGlu 197 | | | |
| DB | 4752 | GGGCCCTCTGGGCTCCAGGTCCACCTGGTGAAGTCAATT-----CAG 4793 | | | |
| QY | 198 | ProThrPro-AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLy 217 | | | |
| DB | 4794 | CCTTTACCAATCT------TGCTCTCCAAAAAACGA----- 4824 | | | |
| QY | 217 | sCysTrpThrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAs 237 | | | |
| DB | 4825 | -----GAAGACATACTGAAGCATGCAAG-- 4848 | | | |
| QY | 237 | pLysSerSerHisLeuValPheIleAsnThrArgGluGluGln----- 251 | | | |
| DB | | ----- 251 | | | |

| | | | | | |
|------------------------|----------|---------------|------|--|--|
| Alignment Scores: | | | | | |
| Pred. No.: | 3,35e-16 | Length: | 6158 | | |
| Score: | 453.00 | Matches: | 130 | | |
| Percent Similarity: | 39.81% | Conservative: | 38 | | |
| Best Local Similarity: | 30.81% | Mismatches: | 76 | | |

US-09-954-456-762
 ; Sequence 762, Application US/09954456
 ; Patent No. US20020115057A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Young, Paul
 ; TITLE OF INVENTION: Sets
 ; FILE REFERENCE: 689290-76
 ; CURRENT APPLICATION NUMBER: US/09/954,456
 ; CURRENT FILING DATE: 2001-09-18
 ; PRIOR APPLICATION NUMBER: US/60/233,617
 ; PRIOR FILING DATE: 2000-09-18
 ; PRIOR APPLICATION NUMBER: US/60/234,052
 ; PRIOR FILING DATE: 2000-09-20
 ; PRIOR APPLICATION NUMBER: US/60/234,923
 ; PRIOR FILING DATE: 2000-09-25
 ; PRIOR APPLICATION NUMBER: US/60/235,134
 ; PRIOR FILING DATE: 2000-09-25
 ; PRIOR APPLICATION NUMBER: US/60/235,637
 ; PRIOR FILING DATE: 2000-09-26
 ; PRIOR APPLICATION NUMBER: US/60/235,638
 ; PRIOR FILING DATE: 2000-09-26
 ; PRIOR APPLICATION NUMBER: US/60/235,711
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: US/60/235,720
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: US/60/235,840
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: US/60/235,863
 ; PRIOR FILING DATE: 2000-09-27
 ; NUMBER OF SEQ ID NOS: 2276
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 762
 ; LENGTH: 6158
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: n=a,t,g or c
 ; US-09-954-456-762

Query Match: 23.35% Indels: 180
DB: 10 Gaps: 18

US-09-763-712a-2_COPY_206_547 (1-342) x US-09-954-456-762 (1-6158)

QY 43 GlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 62
DB 4032 GGACCTCCAGGTGCCAAGGGGCCAGGTGATGATGGCCCTAAGGTTAACCCGGGTCT 4091

QY 63 Thr----- 63
DB 4092 GTTGGTTTCCGGAGATCCTGGTCTCTCTGGGAACTTGGCCCTGCAGGTCAAGATGT 4151

QY 64 -----GlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAla 81
DB 4152 GTTGGTGGTGCACAAAGGTGAAGATCGAGATCTGTGTCAACCGGTCTCTGGCCCATCT 4211

QY 82 GlyGluArgGlyProIleGlyPro-----AlaGlyProProGlyGluArgGlyGly 98
DB 4212 GGTGAGCTGGCCACACAGGTCTCTCTGGAACACGAGGTCTCTCTGGAGCTGCAGGTGCA 4271

QY 99 LysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysPro--- 117
DB 4272 GAGGGAAGACAAGGTGAAGAGTCTAAGGGGGAAGCAGGTGCAGAGGTCTCTCTGGA 4331

QY 118 -----GlyProGlnGlyProSerGlyAspProGlyPro----- 128
DB 4332 AAAACCGCCAGTCCGTCTCAGGACCTGCAGGAACCTGGTCCAGAAAGTCTTCGG 4391

QY 129 -----ProGlyProProGlyLysGluGlyLeu----- 137
DB 4392 GGCATCCCTGGTCTGTGGAGAACAAAGGTCTCCCTGGAGCTGCAGGCCAAGATGCACCA 4451

QY 138 ProGlyProGlnGlyProGlyPhe----- 146
DB 4452 CTTGGTCTATGGACCTCTCGCTTACCTACCTGTCTCAAAGGTGACCTGGCTCCAAAGGT 4511

QY 146 ----- 146
DB 4512 GAAAGGGACATCCTGGTTTAATTGGCCCTGATTGGTCTCCTCAGGAGAACAGGGGAAAAA 4571

QY 147 -----GlnGlyLeuGlnGlyThrValGlyGluPro-----GlyVal 158
DB 4572 GTGTACCGAGGCTCCTCGAATCTCAAGGATCTCCAGGAGGAAAAGGGGATGGGGAAT 4631

QY 159 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys--- 177
DB 4632 CTTGGTCTGTGGTCCCTTAGTCCACCTGGTCTCTCCAGGTACCAGGTCTCTCAAGGC 4691

QY 177 ----- 177
DB 4692 CCAAGGGTAACAAAGGCTCTACTGGACCCGCTGGCCAGAAAGGTGCACAGTGTCTTCCA 4751

QY 178 GlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGlu 197
DB 4752 GGGCTCTCTGGCCCTCCAGGTCCACCTGGTGAATCATP-----CAG 4793

QY 198 ProThrPro-AlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAspLys 217
DB 4794 CCTTACCAATCT-----TGTCTCTCAAAAAACGA----- 4824

QY 217 sCysTrpTrpPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAs 237
DB 4825 -----GAAGACATACTGAAGGCATGAAG-- 4848

QY 237 LysSerSerHisLeuValPheIleAsnThrArgGluGluGln----- 251
DB 4849 -CAGATGCAGATGATAATATTCTTGATTACTCGGATGGAAGAAATATTGGTTCC 4907

QY 252 -----GlnTrpIleLysLysGlnMe 258
DB 4908 CTCATTTCCCTGAACACACATCAGCATATGAAATTTCCATTTGGTACTCAGACCAAT 4967

QY 258 tValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLys 278
DB 4968 CCA-GCCCGAAGTGTAAAGACCTGCAACTCAGCATCTGACTTCCCAGATGGTGAATA 5026

QY 278 strPleuasp-----GlyThrSerProasp-----TyrLysAsnTrpLys 291
DB 5027 TTGGATGTGATCTAACCAAGGTGTCTCAGGAGATCTCTCAAGATTTACTGTAATTTTCA 5086

QY 291 salAGlyGlnProAspAsnTrpGlyHisGlyProGlyGluAspCysAlaGlyLe 311
DB 5087 ATCTGGT-----GGTGAGACTTGC----- 5105

QY 311 uIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysG 331
DB 5106 -ATTTATCCA-----GACAAAAAATCTGAGGGAGTAAG-AATTTCTATCATGCG 5151

QY 331 ulys 332
DB 5152 CAAA 5155

RESULT 7
US-09-954-456-782
; Sequence 782, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 782
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-782

Alignment Scores:
Pred. No.: 3.83e-16 Length: 6728
Score: 452.50 Matches: 99
Percent Similarity: 52.11% Conservative: 12
Best Local Similarity: 46.48% Mismatches: 53
Query Match: 23.32% Indels: 49
DB: 10 Gaps: 8

US-09-763-712a-2_COPY_206_547 (1-342) x US-09-954-456-782 (1-6728)

QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
DB 1485 GTTCAAGGACCCCTCTGGCCCTCTGGAGGAAGGAACCCAGGAGCTCGAGGTGAACCC 1544

QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro----- 75
Db 1545 GGACCCACTGGCTGCCCGACCCCTGGCGAGCTGGTGGACCTGGTAGCGTGGTTTC 1604

QY 76 -----GlyProGlyProAlaGlyGluArgGlyProIleGly 88
Db 1605 CTTGCGGAGATGGTGGTGGTCCCAAGGCTCCGCTGGTGAAGCTGGTTCTCCTGGC 1664

QY 89 ProAlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGlySerGln 105
Db 1665 CCGCTGGCCCCAAGGATCTCTGCTGAAGCTGGTCCGCTGAAGCTGGTGGTTC 1724

QY 106 GlyProLysGlySerArgLysProGlyLysProGlyProGlnGlyProSerGlyAsp 125
Db 1725 GGTGCAAGGGTCTGACTGGAAGCCCTGGCAGCCCTGGTCTGATGGCAAACT----- 1778

QY 126 GlyProGlyProGlyProGlyLysGlyGluLysGlyLysGlyProGlnGlyProPro 144
Db 1779 ---GGCCCCCTGGTCCCGCGCTCAAGATGTCGCCCGGACCCCGGACCCCGGACCCCTGGT 1835

QY 145 -----GlyPheGlnGlyLeuGlnGlyThrValGlyGluPro 156
Db 1836 GCCCTGGTCAAGCTGGTGTGATGGATTCCTGGACCTAAAGGTGCTGCTGGAGGCC 1895

QY 157 GlyValProGlyProArgGlyLeuProGlyLeuProGlyVal----- 170
Db 1896 GCAAGGCTGGAGAGGAGTTCCTGGACCCCTGGTGGTCCGCTGGCGAGAGGT 2015

QY 171 ProGlyMetProGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 187
Db 1956 GATGAGAGGCTGGAGCTCAGGAGCCCTGGCCCTGGTGGTCCGCTGGCGAGAGGT 2015

QY 188 -----AlaValProLeuAlaLeuGlnAsnGluProThrProAla 201
Db 2016 GAACAAGGCCCTGGTGGTCCCTGGTGGTCCCTGGTGGTCCCTGGTGGTCCCT 2072

QY 202 -----ProGluAspAsnGlyCysPro 208
Db 2073 CCAGGTGAAGCAGGCAACCTGGTGAACAGGCTGGTTCCT 2111

RESULT 8
US-09-880-107-3946
; Sequence 3946, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darcil T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3946
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 274615
US-09-880-107-3946

Alignment Scores:
Pred. No.: 3.83e-16 Length: 6728
Score: 452.50 Matches: 99
Percent Similarity: 52.11% Conservative: 12
Best Local Similarity: 46.48% Mismatches: 53
Query Match: 23.32% Indels: 49

Db: 10 Gaps: 8
US-09-763-712a-2_COPY_206_547 (1-342) x US-09-880-107-3946 (1-6728)

QY 41 LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
Db 1485 GTTCAAGGACCCCTGGCCCTGCTGGAGAGGAAGAAAGCAGGAGCTCGAGGTGAACCC 1544

QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluPro----- 75
Db 1545 GGACCCACTGGCTGCCCGACCCCTGGCGAGCTGGTGGACCTGGTAGCGTGGTTTC 1604

QY 76 -----GlyProGlyProAlaGlyGluArgGlyProIleGly 88
Db 1605 CTTGCGGAGATGGTGGTGGTCCCAAGGCTCCGCTGGTGAAGCTGGTTCTCCTGGC 1664

QY 89 ProAlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGlySerGln 105
Db 1665 CCGCTGGCCCCAAGGATCTCTGCTGAAGCTGGTCCGCTGAAGCTGGTGGTTC 1724

QY 106 GlyProLysGlySerArgLysProGlyLysProGlyProGlnGlyProSerGlyAsp 125
Db 1725 GGTGCAAGGGTCTGACTGGAAGCCCTGGCAGCCCTGGTCTGATGGCAAACT----- 1778

QY 126 GlyProGlyProGlyProGlyLysGlyGluLysGlyLysGlyProGlnGlyProPro 144
Db 1779 ---GGCCCCCTGGTCCCGCGCTCAAGATGTCGCCCGGACCCCGGACCCCGGACCCCTGGT 1835

QY 145 -----GlyPheGlnGlyLeuGlnGlyThrValGlyGluPro 156
Db 1836 GCCCTGGTCAAGCTGGTGTGATGGATTCCTGGACCTAAAGGTGCTGCTGGAGGCC 1895

QY 157 GlyValProGlyProArgGlyLeuProGlyLeuProGlyVal----- 170
Db 1896 GCAAGGCTGGAGAGGAGTTCCTGGACCCCTGGTGGTCCGCTGGCGAGAGGT 1955

QY 171 ProGlyMetProGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 187
Db 1956 GATGAGAGGCTGGAGCTCAGGAGCCCTGGCCCTGGTGGTCCGCTGGCGAGAGGT 2015

QY 188 -----AlaValProLeuAlaLeuGlnAsnGluProThrProAla 201
Db 2016 GAACAAGGCCCTGGTGGTCCCTGGTGGTCCCTGGTGGTCCCTGGTGGTCCCT 2072

QY 202 -----ProGluAspAsnGlyCysPro 208
Db 2073 CCAGGTGAAGCAGGCAACCTGGTGAACAGGCTGGTTCCT 2111

RESULT 9
US-09-954-456-786
; Sequence 786, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27

[illegible]

```

Db      3164   ATTTGCGGGCCCTCCTGGGGCCCCGTGGTCTCCTGTGTGTGGTAGTCCTGGAGTCAAC 3233
          :::                |||                               :::|||       |||
Qy      206 GlyCysPro 208
          ||| |||
Db      3224   GGTGCTCTCT 3232
          ||| |||

RESULT 10
US-09-880-107-2094
; Sequence 2094, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2094
; LENGTH: 5416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 J03464
US-09-880-107-2094

Alignment Scores:
Pred. No.:           9,05e-16              Length:           5416
Score:             444.00                 Matches:            107
Percent Similarity: 47.15%               Conservative:       17
Best Local Similarity: 40.68%             Mismatches:         64
Query Match:        22.89%                Indels:             76
DB:                  10                   Gaps:               9

US-09-763-712A-2_COPY_206_547 (1-342) x US-09-880-107-2094 (1-5416)

Qy      20 IleMetGluGlyMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThr 39
          :::::  ||:::||:::||:::||:::||::|||
Db      2448 CTGGTCTCACAGGTGAATGTGTAAACCCTGGCAGAGATGGTGCTCGTGGTG---CTCATG 2504
          :::::  ||:::||:::||:::||:::||::|||

Qy      40 IleLeuGlnGlyProGlyProArgGlyProArgGlyAspArg-----GlySer 56
          :::::  ||| ||||| ||||| ||||| ||||| |||||
Db      2505 GTGCTGTGA-GGTGCCCTGGTCTCTGTGTGGACCACAGGTGACCGGGCGCAAAGCTGGGGCT 2563
          :::::  ||||| ||||| ||||| ||||| ||||| |||||

Qy      57 GlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGly 76
          ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2564 GCTGTGCTCTGCTGCTCTGCTGCTCGGGAAGCCCTGGTGAACGTGGCGAGGTGGT 2623
          ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy      77 ProProGlyProAlaGlyGlnArgGlyProIleGlyProAlaGlyProPro----- 93
          ||| ||||| ||||| ||||| ||||| ||||| |||||

Db      2624 CTTGCTGGCCCCAACAGGATTGCTGTCCGGCTGGTGTGCTGTGCTCAACCCGGGTGCTAAA 2683
          ||| ||||| ||||| ||||| ||||| ||||| |||||

Qy      94 GlyGluArgGlyGlyLysGlySerLysGlySerGln----- 105
          ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db      2684 GGAGAAAGAGAGCAAGGGCCTAAGGCTGAAACGGTGTGTTGGTGCACACAGGCCCC 2743
          ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy      106 -----GlyProLysGlySerArgGly 112
          ||||| ||||| ||||| ||||| ||||| |||||

Db      2744 GTTGGAGCTGTCGCCCCAGCTGCTCCAAATGTTGCCCGGCTCTGTGCTGSGAAGTCGTGGT 2803
          ||||| ||||| ||||| ||||| ||||| |||||

Qy      113 Ser-----ProGlyLysPro 117
          ||||| ||||| ||||| |||||

Db      2804 GATGGAGGCCCCCCTGGTATGACTGTTTCCTGTGTGCTGTGACGAGCTGGTCCCCCA 2863
          ||||| ||||| ||||| ||||| ||||| |||||

Qy      118 GlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyProGlyLysGlyLeu 137
          ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


QY 65 AsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArg 84
||||| ||||||| ||| |||||| ||||||| |||
Db 1427 GCCAAGGGCCCTAAGGGTGAAACGGTGTGTGGTCCACAGGCCCGGTGGAGTGCT 1486

US-09-763-712A-2_COPY_206_547 (1-342) x US-

0.3 0.3

0.3 0.3

```
Db 2246 GGTCTGCTGGTCTCGGGAAGCCCTGGTGAACGTGGCGAGGTGGTCTGCTGCGCC 2305
QY 63 Thr-----Gly 64
Db 2306 AACGGATTTCGTGGTCCGCTGGTCTGCTGCTCAACCGGTCTAAAGAGAGAGAGA 2365
QY 65 AsnLysGlyGlnLysGlyGluProGlyProGlyProGlyProGlyProGlyGluArg 84
Db 2366 GCCAAGAGCCCTAAGGCTGAACGGTGTGTGTGTCACAGGCCCTGTTGAGAGTGT 2425
QY 85 GlyProIleGlyProAlaGlyProGlyGluArgGlyGlyGlySerLysGlySer 104
Db 2426 GGCCCCAGCTGGTCCCAATAGTCCCGCTGCTGCTGAAGTCGTGGTGTATGAGGCC 2485
QY 105 GlnGlyProLysSerArgGlySerProGlyLysProGlyProGlyProGlyProGly 122
Db 2486 CCTGGTATGACTGGTTTCCTGCTGCTGGAGCGACTGGTCCCGCCAGGACCTCTGT 2545
QY 123 ---SerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyPro 141
Db 2546 ATTCTGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2605
QY 142 -----GlyProProGlyPhe 146
Db 2606 GTGACCAAGGTCAGTTGGCCGAAGTGAAGTAGTGGTGGTGGTGGTGGTGGTGG 2665
QY 147 GlnGlyLeuGln-----GlyThrValGlyGluProGlyValProGly 160
Db 2666 GCTGGTGAGAGGGTCCCTCTGGAGAGGCTGGTACTGCTGGACCTCTGGCCTCAG 2725
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyPro 180
Db 2726 CCTCAGGGCTCTTCTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2785
QY 181 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 197
Db 2786 GGCTACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2845
QY 198 Pro-----ThrProAlaProGluAspAsnGlyCysPro 208
Db 2846 CTGGGGCCCGTGGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2902

RESULT 13
US-09-925-299-206
; Sequence 206, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 206
; LENGTH: 5145
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (5126)
; OTHER INFORMATION: n equals a,t,g, or c
```

```
; NAME/KEY: misc_feature
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Alignment Scores:
Pred. No.: 1.33e-15 Length: 5145
Score: 440.50 Matches: 95
Percent Similarity: 49.32% Conservative: 13
Best Local Similarity: 43.38% Mismatches: 58
Query Match: 22.71% Indels: 53
DB: 9 Gaps: 6

US-09-763-712a-2_copy_206_547 (1-342) x US-09-925-299-206 (1-5145)
QY 43 GlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 62
Db 2248 GGTCTGCTGGTCTCGGGAAGCCCTGGTGAACGTGGTGAAGTGGTGGTGGTGGT 2307
QY 63 Thr-----Gly 64
Db 2308 AATGGATTTCGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2367
QY 65 AsnLysGlyGlnLysGlyGluProGlyProGlyProGlyProGlyProGlyProGly 84
Db 2368 GCCAAGAGCCCTAAGGCTGAACGGTGTGTGTGTCACAGGCCCTGTTGAGAGTGT 2427
QY 85 GlyProIleGlyProAlaGlyProGlyGluArgGlyGlyGlySerLysGlySer 104
Db 2428 GCCCCAGCTGGTCCCAATAGTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2487
QY 105 GlnGlyProLysSerArgGlySerProGlyLysProGlyProGlyProGlyProG 122
Db 2488 CCTGGTATGACTGGTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2547
QY 123 ---SerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGly 141
Db 2548 ATTCTGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2607
QY 142 -----GlyProProGlyPhe 146
Db 2608 GTGACCAAGGTCAGTTGGCCGAAGTGAAGTAGTGGTGGTGGTGGTGGTGGTGG 2667
QY 147 GlnGlyLeuGln-----GlyThrValGlyGluProGlyValProGly 160
Db 2668 GCTGGTGAGAGGGTCCCTCTGGAGAGGCTGGTACTGCTGGACCTCTGGCCTCAG 2727
QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyPro 180
Db 2728 CCTCAGGGCTCTTCTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2787
QY 181 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 197
Db 2788 GGCTACCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2847
QY 198 Pro-----ThrProAlaProGluAspAsnGlyCysPro 208
Db 2848 CTGGGGCCCGTGGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2904

RESULT 14
US-09-925-299-206
; Sequence 206, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
```


QY 198 Pro-----ThrProAlaProGluAspAsnGlyCysPro 208
||| :||| ||||| |||
Db 3179 CCTGGGGCCGGTCTCTCTGGTGTGGGTAGTCTGAGTCAACGGTGCTCCT 3235

Search completed: March 21, 2003, 12:32:28
Job time : 186.172 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:08:41 ; Search time 1130.43 Seconds
(without alignments)
4899.781 Million cell updates/sec

Title: US-09-763-712A-2_COPY_206_547

Perfect score: 1940

Sequence: 1 MQQDLMSRLDTEVANLSVI.....EDVNNFCKDKRETLYSSAL 342

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09763712/runat_14032003_100950_18109/app_query.fasta_1.1877
-DB=BST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712 -CGN_1_1_4575=runat_14032003_100950_18109 -NCPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_rrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|--------------------|
| 1 | 1801 | 92.8 | 3305 | 11 | BC009162 Mus muscu |
| 2 | 1256 | 64.7 | 861 | 14 | BQ713873 AGENCOURT |
| 3 | 1061 | 54.7 | 906 | 13 | BI456109 603172765 |
| 4 | 1022 | 52.7 | 580 | 12 | BE910803 601661855 |
| 5 | 1007 | 51.9 | 668 | 14 | BQ674807 AGENCOURT |
| 6 | 957 | 49.3 | 601 | 10 | BE290299 601089246 |
| 7 | 937 | 48.3 | 957 | 14 | BQ955927 AGENCOURT |
| 8 | 916.5 | 47.2 | 967 | 9 | AL568743 AL568743 |
| 9 | 842 | 43.4 | 552 | 10 | AW958053 EST370123 |
| 10 | 831 | 42.8 | 638 | 10 | BB248064 BB248064 |
| 11 | 767 | 39.5 | 808 | 14 | BQ771366 UI-M-F10- |
| 12 | 696.5 | 35.9 | 936 | 14 | BQ891432 AGENCOURT |
| 13 | 669 | 34.5 | 500 | 14 | BM676508 UI-E-E30- |
| 14 | 663 | 34.2 | 507 | 14 | BM713891 UI-E-E30- |
| 15 | 627.5 | 32.3 | 591 | 13 | BI445884 dal33d10. |
| 16 | 600 | 30.9 | 357 | 10 | AW379436 CM4-HT024 |
| 17 | 594 | 30.6 | 350 | 9 | AA012704 RPU1402CG |
| 18 | 588 | 30.3 | 339 | 9 | AA304251 EST17006 |
| 19 | 560 | 28.9 | 517 | 9 | AL543000 AL543000 |
| 20 | 537 | 27.7 | 310 | 9 | AA361740 EST71069 |
| 21 | 478 | 24.6 | 943 | 14 | BQ887163 AGENCOURT |
| 22 | 460 | 23.7 | 441 | 10 | BB749743 BB749743 |
| 23 | 459.5 | 23.7 | 1082 | 14 | BM907108 AGENCOURT |
| 24 | 458.5 | 23.6 | 569 | 10 | AV616076 AV616076 |
| 25 | 456.5 | 23.5 | 792 | 13 | BI456015 603170460 |
| 26 | 456 | 23.5 | 420 | 12 | BE929813 RC3-GN004 |
| 27 | 455.5 | 23.5 | 896 | 14 | BQ955616 AGENCOURT |
| 28 | 453.5 | 23.4 | 877 | 12 | BQ681943 602629884 |
| 29 | 452 | 23.3 | 688 | 13 | BI249235 602995511 |
| 30 | 452 | 23.3 | 3951 | 11 | BC013626 Mus muscu |
| 31 | 451.5 | 23.3 | 921 | 14 | BQ921102 AGENCOURT |
| 32 | 451.5 | 23.3 | 1192 | 14 | BM905999 AGENCOURT |
| 33 | 449.5 | 23.2 | 907 | 14 | BQ222937 AGENCOURT |
| 34 | 448 | 23.1 | 946 | 14 | BQ942234 AGENCOURT |
| 35 | 447.5 | 23.1 | 1029 | 17 | CNS04VYG Tetraodon |
| 36 | 446.5 | 23.0 | 608 | 13 | BM425818 pgf2c.pk0 |
| 37 | 446 | 23.0 | 651 | 10 | BB613976 BB613976 |
| 38 | 446 | 23.0 | 819 | 12 | BG679800 602626481 |
| 39 | 446 | 23.0 | 1120 | 14 | BM809933 AGENCOURT |
| 40 | 445.5 | 23.0 | 1014 | 14 | BQ942135 AGENCOURT |
| 41 | 445 | 22.9 | 972 | 14 | BQ896478 AGENCOURT |
| 42 | 444.5 | 22.9 | 678 | 13 | BM486385 pgmlc.pk0 |
| 43 | 444.5 | 22.9 | 781 | 13 | BI454661 603170558 |
| 44 | 444 | 22.9 | 429 | 12 | BE929798 RC3-GN004 |
| 45 | 444 | 22.9 | 991 | 14 | BQ713634 AGENCOURT |

ALIGNMENTS

RESULT 1
BC009162
LOCUS BC009162 3305 bp mRNA linear HTC 07-AUG-2002
DEFINITION Mus musculus, Similar to scavenger receptor with C-type lectin,
clone IMAGE:2811487, mRNA.
ACCESSION BC009162
VERSION BC009162.1 GI:14714370
KEYWORDS HTC.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3305)
AUTHORS Strausberg, R.
TITLE Direct Submission

JOURNAL

Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: Lother Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,

Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,

Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 5 Row: h Column: 10

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF analysis

This clone has the following problem: incomplete processing.

FEATURES

source

1..3305
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="IMAGE:2811487"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."
/clone_lib="NCL CGAP_Mam5"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

BASE COUNT 1031 a 757 c 760 g 757 t
ORIGIN

Alignment Scores:

Pred. No.: 2,02e-108 Length: 3305
Score: 1801.00 Matches: 313
Percent Similarity: 95.91% Conservative: 15
Best Local Similarity: 91.52% Mismatches: 14
Query Match: 92.84% Indels: 0
DB: 11 Gaps: 0

US-09-763-712A-2_COPY_206_547 (1-342) x BC009162 (1-3305)

QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
|||||
DB 1296 ATCCAGCAAGACATGATGAGGTCAAGTATGACACTGAAGTGGCCAACTATCAGTGGIT 1355
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
|||||
DB 1356 ATGGAAGAGATGAACCTGGTGTACTCCAGACGGTTCAGTTCATCAAGAACTTTACCAIT 1415
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgSerGlnGlyProPro 60
|||||
DB 1416 CTACAAGTCTCTCTGGCCCCAGAGGTCCAAAGGTGCAGAGAGATCTCAGGACCACCT 1475
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
|||||
DB 1476 GGTCCAACGTGGCAACAAGGACAGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1535
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyGlyGly 100
|||||
DB 1536 GCGGGGTGAGAGGGGCAACAATTGGACCAAGTCCCTCCCTGGAGAGCGTGGCAGCAAGGA 1595
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyProGlyProGln 120
|||||
DB 1596 TCCAAAGGCTCAGAGGGTCCCAAGAGATCTCTGTGGTCCCCAGGGAAGCGTGGCCCTCAA 1655

QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyPro 140
|||||
DB 1656 GGACCTAGTGGGACCCAGGACCCAGGACCCAGGACCCAGGACCCAGGACCCAGGACCCAGGACCC 1715
QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
|||||
DB 1716 CAGGGCCCTCTGGCTTCCAGGGACTACAGGGCAGCTGTGGGTGAGCTGTGGGTGAGCTGTGG 1775
QY 161 ProArgGlyLeuProGlyProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
|||||
DB 1776 CTTGGGGGTGGCAGGCTTGGCAGGGGTGCCAGGAGTGGCTGGGCTTAAGGACCACCT 1835
QY 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
|||||
DB 1836 GGGCCCTCCAGGGCCCTCAGGAGCAATGGAGCCATTGGCTCGCAGATGAACCAACCCCA 1895
QY 201 AlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAspLysCysTyrTyr 220
|||||
DB 1896 GCATCAGAGGTCAACGGATGTCCACCTCACTGGAAGAACTTCACAGATAAATGCTACTAT 1955
QY 221 PheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSer 240
|||||
DB 1956 TTTTCATTGGAAAAAGAAATTTTGAAGATGCTAAAGCTTTTCTGTGAAGACAAATCTTC 2015
QY 241 HisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGly 260
|||||
DB 2016 CATCTCGTTTCATAAATCAAGAGAAGAAGCAAGCAATGGATAAAGGATACCCGTGGG 2075
QY 261 ArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeu 280
|||||
DB 2076 AGAGAAAGCAATTTGGATCGCCCTCACAGACTCAGAAAGCAAGAAAGCAAGCAATTTGG 2135
QY 281 AspGlyThrSerProAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHis 300
|||||
DB 2136 GACGGGTACCTGTTGATTACAAAACTGGAAAGCTGGACCAACCACTAACTACTGGGCACT 2195
QY 301 GlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPhe 320
|||||
DB 2196 GGCCATGGCCAGGAGAGACGTGTGCTGGCTTACGACGAGACAGTGGATGACTTC 2255
QY 321 GlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSer 340
|||||
DB 2256 CAGTGTGATGAATCAATCAATCAATTTGTTGAGAGAAAGGAGGAGGAGGAGGAGGAGGAG 2315
QY 341 AlaLeu 342
|||
DB 2316 ATATTA 2321
RESULT 2
B0713873
LOCUS
DEFINITION
5', mRNA sequence.
AGENCOURT_8482828 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6306240
B0713873
ACCESSION
VERSION
B0713873.1
KEYWORDS
EST.
house mouse.
Mus musculus
ORGANISM
REFERENCE
1 (bases 1 to 861)
AUTHORS
NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: ResGen, Invitrogen Corp
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
 Plate: L1AM13721 row: n column: 01
 High quality sequence stop: 598.

FEATURES

source

1. 861
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:6306240"
 /lab_host="NIH_MGC_129"
 /note="Organ: olfactory epithelium; Vector:
 pCMV-SPORT6.1 ccdb; Site_1: EcoRV; Site_2: NotI; Cloned
 unidirectionally. Primer: Oligo dt. Average insert size
 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
 is a NIH_MGC Library."

BASE COUNT 221 a 253 c 239 g 147 t 1 others

ORIGIN

Alignment Scores:
 Pred. No.: 3.18e-73 Length: 861
 Score: 1256.00 Matches: 229
 Percent Similarity: 91.01% Conservative: 14
 Best Local Similarity: 85.77% Mismatches: 20
 Query Match: 64.74% Indels: 4
 DB: 14 Gaps: 0

US-09-763-712a-2_COPY_206_547 (1-342) x BQ713873 (1-861)

QY 5 LeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMet 24
 DB 13 ATGATGAGTCAAGTGTAGACACTGAAGTGGCCAACTTATCATGTTGTATGAAGAGATG 72
 QY 25 LysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyPro 44
 DB 73 AAACGTGGTTGACTCAAGCAGCGTCAGCTCATCAAGAAGCTTTACCATTCTCAAGGTCT 132
 QY 45 ProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGly 64
 DB 133 CTGCGCCCGAGAGTCCAAAGAGTGCACAGAGATCTCAGGGACCACTGGTCCAACTGGC 192
 QY 65 AsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArg 84
 DB 193 AACAGGGACAAAGGAGAGGAGGAGGAGCTGGTGCACCTGGCCCTCGGGTGGAGAGG 252
 QY 85 GlyProIleGlyProAlaGlyProGlyGluArgGlyGlyGlySerLysGlySer 104
 DB 253 GCCAATTTGGACCACTCGGCCCTCTCTGGAGAGCTGGCAGCAAGGATCCAAAGCTCA 312
 QY 105 GlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGly 124
 DB 313 CAGGGTCCCAAGGATCTGTTGGTCCCCAGGGAAGCCTGGCCCTCAAGGACCTAGTGGG 372
 QY 125 AspProGlyProGlyProGlyProGlyLysGlyGluGlyLeuProGlyProGlnGlyProPro 144
 DB 373 GACCCAGGACCAACAGGTCCACGAGGAGGATGACCTCCCTGGCCCTCAGGGCCCTCT 432
 QY 145 GlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeu 164
 DB 433 GCCTTCCAGGACTACAGGCACTGTGGTGAGCTGGAGTACCTGGACCTCGGGGTG 492
 QY 165 ProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGly 184
 DB 493 CCAGGCTTCCAGGGTGCAGGCATGCTGGGCTCCAGGGAAGCCTAGGGCCCTCCAGGC 552
 QY 185 ProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAsp 204
 DB 553 CCTCAGGAGCAATGAGCCATGTGCTCTGCAAAATGAACCAACCCANCATCAGAGTTC 612
 QY 205 AsnGlyCysProHisThrLysAsnPheThrAspLysCysTyrTyrPheSerVal-Gl 224
 DB 613 AACGATGTCCGCCCTCAGTGGAGAACTTCACAGATATAATGCTACTATTTTCCCTGGCA 672
 QY 224 ulysGluIlePheGluAspAlaLysLeu-PheCysGluAspLysSerHisLeuValP 244

DB 673 AAAAAAATTTTCGAAATGCTTACCTTTTCTCGGAACAAAAATCTCCCATCCGGGT 732
 QY 244 heileAsnThrArgGluGlnGlnTrp-Ile-LysLysGlnMetValGlyArgGluSe 263
 DB 733 TCATAACCCCAAGAAAAACAGCCATGGATATAAAACCTACCTTCGGGGACCAAAAC 792
 QY 263 rHisTrpIleGly 267
 DB 793 CCGCTGGATCGGG 805
 RESULT 3
 BI456109 906 bp mRNA linear EST 21-AUG-2001
 LOCUS 603172765F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5251888 5',
 DEFINITION mRNA sequence.
 ACCESSION BI456109
 VERSION BI456109.1 GI:15246765
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 906)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM1636 row: b column: 17
 High quality sequence stop: 730.
 Location/Qualifiers
 1. 906
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:5251888"
 /clone_lib="NCI_CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

BASE COUNT 247 a 228 c 256 g 175 t

ORIGIN

Alignment Scores:

Pred. No.: 1.97e-60 Length: 906
 Score: 1061.00 Matches: 203
 Percent Similarity: 87.76% Conservative: 12
 Best Local Similarity: 82.86% Mismatches: 26
 Query Match: 54.69% Indels: 6
 DB: 13 Gaps: 2

US-09-763-712a-2_COPY_206_547 (1-342) x BI456109 (1-906)

QY 1 MetGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
 DB 123 ATGCAGCAAGACATGATGAGGTCAAAGTTAGACACTGAAGTGGCCAACTTATCAGTGGT 182
 QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
 DB 183 ATGGAAGAGATGAAACTGGTTGACTCCAAAGCAGCGGTCACTCATCAAGAACTTTACCAT 242

QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
 Db 243 CTCAAGAGTCTCTCGCCAGAGGTCCAAAAGGTGACAGAGGATCTCAGGACCACT 302
 QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyPro 80
 Db 303 GTTCCAACTGGCAACAGAGGACAGAAAGGAGAGAGGAGGCTGGTCCAGCTGGCCCT 362
 QY 81 AlaGlyGluArgGlyProLysGlyProAlaGlyProProGlyGluArgGlyGlyLysGly 100
 Db 363 GCGGTGAGAGGGGACATTTGGACCATCTGCGCCCTCTGGAGAGGCTGGCAGCAAGGA 422
 QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGln 120
 Db 423 TCCAAAGGCTCACAGGTCCTCCAAAGGATCTCGTGGGTCCCAAGCTGGCCCTCAA 482
 QY 121 GlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyPro 140
 Db 483 GGACCTAGTGGGACCCAGGACCAACAGGTCCACCAAGGCAAGATGACTCCCTGGCCCT 542
 QY 141 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
 Db 543 CAGGGCCCTCTGGCTTCAGGACTACAGGCACTGTGGGTGAGCTGAGTACCTGA 602
 QY 161 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 180
 Db 603 CCTCGGGGTTCAGGCTTGCAGGGTCCAGGATGCTGGGCTAAGGACCACT- 661
 QY 181 GlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrPro 200
 Db 662 AGCCCTCCAGGGCCCTCAGGATG- GAGCCATTTGGTCTGAGAAATGAACCAAGCCCA 720
 QY 201 AlaProGlu--AspAsnGlyCysProProHisTriPlysAsnPheThrAsp---LysCysT 219
 Db 721 GCATCAGAGTTCAACGCGATGTCCTCTCACTGGAAAGAACTTTCACGGATTAATGGCT 780
 QY 219 yrrTyrrPheSerVal---GluLysGluIlePheGluAspAlaLysLeuPheCysGluAspL 238
 Db 781 ACTATTTTCCATTCACAAAAGAAATTTTGAAGATGCTTAAGCGTTCTGTGGAGAAA 840
 QY 238 ySerSerHis 241
 Db 841 AATTTTCCAAT 851
 RESULT 4
 BE910803
 LOCUS
 DEFINITION 580 bp mRNA linear EST 29-SEP-2000
 601661855F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3962292 5',
 mRNA sequence.
 ACCESSION BE910803
 VERSION BE910803.1 GI:10407765
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 580)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM9128 row: m column: 13
 High quality sequence stop: 580.
 Location/Qualifiers

source 1..580
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3962292"
 /clone_lib="NCI_CGAP_Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="10 months, virgin"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"
 BASE COUNT 159 a 153 c 161 g 107 t
 ORIGIN
 Alignment Scores: 4.81e-58 Length: 580
 Pred. No.: 1022.00 Matches: 177
 Score: 96.34% Conservative: 7
 Best Local Similarity: 92.67% Mismatches: 7
 Query Match: 52.68% Indels: 0
 DB: 12 Gaps: 0
 US-09-763-712A-2_COPY_206_547 (1-342) x BE910803 (1-580)
 QY 99 LysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGly 118
 Db 6 AAAGATCCAAAGGCTCACAGGTCCTCCAAAGGATCTCGTGGTCCCCAGGAAGCTGCG 65
 QY 119 ProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuPro 138
 Db 66 CCTCAAGGACCTAGTGGGACCCAGGACCAAGGTCCTCCAGGCAAGGATGGACTCCCT 125
 QY 139 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 158
 Db 126 GGCCTTCAGGGCCCTCTCGCTTCAGGACTACAGGCACTGTGGGTGAGCTGGAGTA 185
 QY 159 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 178
 Db 186 CTTGGACCTCGGGGTTCAGGCTTGCAGGGGTGCCAGGATGCTGGCCCTAAGGGA 245
 QY 179 ProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 198
 Db 246 CCACCTGGCCCTCCAGGCTTCAGGACCAATGGAGCCATGGCTCTGAGATGAACCA 305
 QY 199 ThrProAlaProGluAspAsnGlyCysProProHisTriPlysAsnPheThrAspLysCys 218
 Db 306 ACCCAGCATCAGAGGTCAACGGATGTCCGCTCACTGCAAGAACTTCACAGATAATGC 365
 QY 219 TyrTyrrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLys 238
 Db 366 TACTATTTTCATTTGAAAAAGAAATTTTGAAGATGCTTAAGCTTTCTGTGAAGACAAA 425
 QY 239 SerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMet 258
 Db 426 TCTTCCCATCTCGTTTCATAAATCAAGAGAAGAACAAATGGATAAAGCAATACC 485
 QY 259 ValGlyArgGluSerHisTriPleGlyLeuThrAspSerGluArgGluAsnGluTrpLys 278
 Db 486 GTGGGAGAGAAAGCCATTGGATGGCCCTCACAGACTCAGACAGGAAGCAATGAAG 545
 QY 279 TrpLeuAspGlyThrSerProAspTyrrLysAsn 289
 Db 546 TGGCTAGACGGGTCACTCTTTGATTACAAAAAC 578
 RESULT 5
 BQ674807
 LOCUS
 DEFINITION BQ674807 668 bp mRNA linear EST 15-JUL-2002
 AGENCOURT_8035172 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6212672
 5', mRNA sequence.
 ACCESSION BQ674807
 VERSION BQ674807.1 GI:21785641

Db 62 GAGGGGACAAATGGACAGCTCGCGCCCTCTCTGGAGAGCGTGGCAGCAAGATCCAAAGG 121
QY 103 ySerGlnGlyProGlySerArgGlySerPro-GlyLysProGlyProGlnGlyProS 123
|||||
Db 122 CTCACAGGTCCTCCAAAGGATCTCGTGGTCCCAGGCTGAGCTGGCCCTCAAGGACCTA 181
QY 123 erGlyaspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGlnGlyP 143
|||||
Db 182 GTGGGACCCAGGACCAAGGTCCTGAGGTCACAGGCAAGGATGGACTCCCTGGCCCTCAGGGCC 241
QY 143 roProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArg 163
|||||
Db 242 CTCCTGCTCCAGGACTACAGGGCACTGTGGTGGAGCTGGAGTACCTGGACCTCGGG 301
QY 163 lyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProP 183
|||||
Db 302 GGTTCAGGCTTGGCAGGGTG-CCAGGCATGCTGGGCTTAAAGGACCACTTGGCCCTC 360
QY 183 roGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProG 203
|||||
Db 361 CAGGCCCCCTCAGGAGCAATGGAGCCATTTGGCTCTGCAGAAATGAACCAACCCAGCATCAG 420
QY 203 luAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTrpPheSerV 223
|||||
Db 421 AGGTCACGGATGTCGCCCTCTCTGGAAGAACTTCACAGATAAATGCTACTATTTTCAT 480
QY 223 alGlyLysGluLeuPheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuV 243
|||||
Db 481 TGGAAAAAGAAAT-TTTGAAGATGCTAAGCTTTTCTGTGAAGACAAATCTTCCCATCTCG 539
QY 243 alPheLeuAsnThrArgGluGluGln-GlnTrpIleLysLysGlnMetValGlyArgGlu 262
|||||
Db 540 TTTTCAATAACTCAAGAGAAGAACAGCCCAATGGTAAAAAAGCATAC-GTGGGGAGAGAA 598
QY 263 Ser 263
:::
Db 599 GCC 601

RESULT 7
BQ955927
LOCUS
DEFINITION AGENCOURT_8763247 NIH_MGC_129 Mus musculus cdna clone IMAGE:6313500
5', mRNA sequence.
BQ955927
VERSION BQ955927.1 GI:22371405
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999).
Contact: Robert Strausberg, Ph.D.
Email: cgapps-r@mail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
CDNA Library Preparation: ResGen, Invitrogen Corp.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13740 row: 1 column: 13
High quality sequence stop: 665.
Location/Qualifiers
i. .957
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6313500"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector:

FEATURES
source

pcMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dT. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH_MGC Library.*
BASE COUNT 260 a 269 c 258 g 169 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 2,73e-52 Length: 957
Score: 937.00 Matches: 182
Percent Similarity: 91.67% Conservative: 5
Best Local Similarity: 89.22% Mismatches: 11
Query Match: 48.30% Indels: 7
DB: 14 Gaps: 2
US-09-763-712A-2_COPY_206_547 (1-342) x BQ955927 (1-957)
QY 1 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIle 20
|||||
Db 353 ATGCAGCAAGACATGATGAGGTCAAAAGTTAGACACTGAAGTGGCCAACTTATCAGTGGTT 412
QY 21 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIle 40
|||||
Db 413 ATGGAAGACATGAAACTGGTTGACTCCAAGCAGCGTCACTCATCAAGAAGACTTTTACCAT 472
QY 41 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 60
|||||
Db 473 CTACAAGGTCTCTCTGGCCCCAGAGGTCCAAAAGGTGACAGAGGATCTCAGGACCACCT 532
QY 61 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 80
|||||
Db 533 GTTCCAACTGGCAACAAGGACAGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 592
QY 81 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyProGlyGluArgGlyLysGly 100
|||||
Db 593 GCGGGTGAGAGGGGGACAAATTGGACCAGTCCGCCCTCTCTGGAGAGCGTGGCAGCAAGGA 652
QY 101 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 120
|||||
Db 653 TCCAAAGGCTCACAGGCTCCCAAGGATCTCGTGGTGGTCCCAAGGAGAGAGAGAGAGAGAG 712
QY 121 GlyProSerGlyAspProGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 140
|||||
Db 713 GGACCTAGTGGGACCCAGGACCAAGGATCCACAGGCAAGGATGGACTTCCCTGGCCCT 772
QY 141 GlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 160
|||||
Db 773 CANGGCCCTCTCGGCTTCAG-GGACTACAGGGCACTGTGGGTGAGCGCTGGAGTACCTGGA 831
QY 161 ProArgGlyLeu-ProGlyLeuProGlyValProGlyMetPro-GlyProLysGlyPro 179
|||||
Db 832 CCTCGGGGGTGGCCAGGCTTGCAGGGGTGCCAGGCATGCCCTGGGGCTTAAGGACCA 891
QY 180 Pro-GlyProProGlyPro---SerGlyAlaValValProLeuAla---LeuGlnAsnG 197
|||||
Db 892 CCTGGGCCCTCCAAAGGCCCTTCAGGAGACAAAGTGGAGCCACTTGGCCTCTCTCGAAGATGA 951
QY 197 uPro 198
|||||
Db 952 ACCA 955
RESULT 8
AL568743
LOCUS AL568743 LTI_FL002_PL1 Homo sapiens cdna clone CS0DE005YH04 3 prime
DEFINITION AL568743 , mRNA sequence.
ACCESSION AL568743
VERSION AL568743.1 GI:12923387
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 967)
Li.W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNAL
COMMENT
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
source
1. 967
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DE005YH04"
/clone_lib="LTI_FL002_PL1"
/lab_host="DH10B"
/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 222 a 221 c 217 g 290 t 17 others
ORIGIN
Alignment Scores:
Pred. No.: 6.06e-51 Length: 967
Score: 916.50 Matches: 174
Percent Similarity: 91.15% Conservative: 1
Best Local Similarity: 90.62% Mismatches: 14
Query Match: 47.24% Indels: 5
DB: 9 Gaps: 3
US-09-763-712A-2_COPY_206_547 (1-342) x AL568743 (1-967)
QY 151 GlyThrValGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyVal 170
DB 947 GCACCGTGGAGCGCTGGTCTCT---GGACCTCGGAGTGCAG---CTTGCTGGGTA 894
QY 171 ProGlyMetProGlyProGlyProGlyProGlyProGlyProSerGlyAlaValVal 190
DB 893 CCAGGCATGCCAGSCCAAG---CCCCCGCCCTCTCTGC-CCATCAGGACGGGTG 838
QY 191 ProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProHis 210
DB 837 CCCCTGSCCTGCAGAAATGAGCAACCCCGGACCGGAGGACAATGGCTGCCCTCAC 778
QY 211 TrpLysAsnPhetheAspLysCysTyrTyrrPheSerValGluLysGluPheGluAsp 230
DB 777 TGAAGAAGCTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTT 718
QY 231 AlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGlu 250
DB 717 GCAAGAGCTTTCTGTGAAGACAAAGCTTCACATCTTGTGTTTATAAACAAGAG 658
QY 251 GlnGlnTrpIleLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAsp 270
DB 657 CAGCAATGATGATAAAGACAGATGTAGGAGAGAGAGCCACTGGATCGCCCTCACAG 598
QY 271 SerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrp 290
DB 597 TCAGAGCGTGAATGAATGAAGTGGTGGATGGAGATCTCCAGACTACAAAATGG 538
QY 291 LysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGly 310
DB 537 AAAGCTGGACAGCCGGAACTGGGGTCTATGC-CATGGCCAGGAGAGACTGTCTGG 479
QY 311 LeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPhelleCys 330
DB 478 TTGATTTATGCTGGCAGTGGGAACGATTCCCAATGTGAAGACGTCAATAACTTCATT 419

QY 331 GluLysAspArgGluThrValLeuSerSerAlaLeu 342
DB 418 GAAAGACAGGAGACAGACTGTCTCATCTGCATTA 383
RESULT 9
AW958053 552 bp mRNA linear EST 01-JUN-2000
LOCUS EST370123 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
DEFINITION AW958053
ACCESSION AW958053
VERSION AW958053.1 GI:8147736
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 115
Seq primer: Reverse.
Location/Qualifiers
source
1. 552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGE"
/note="Vector: pBluescriptSKm"
BASE COUNT 176 a 112 c 143 g 121 t
ORIGIN
Alignment Scores:
Pred. No.: 2.84e-46 Length: 552
Score: 842.00 Matches: 149
Percent Similarity: 98.68% Conservative: 1
Best Local Similarity: 98.03% Mismatches: 1
Query Match: 43.40% Indels: 1
DB: 10 Gaps: 0
US-09-763-712A-2_COPY_206_547 (1-342) x AW958053 (1-552)
QY 192 LeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProHisTrp 211
DB 1 CTGGCCCTGCACAATGAGCAACCCCGGACCGGAGGACAATGGCTGCCCTCACTGG 60
QY 212 LysAsnPhetheAspLysCysTyrTyrrPheSerValGluLysGluPheGluAspAla 231
DB 61 AAGAAGCTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAAGAAATTTT 120
QY 232 LysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGln 251
DB 121 AAGCTTTTCTGTGAGACAAAGCTTCACATCTTGTGTTTATAAACAAGAGAG 180
QY 252 GlnTrpIleLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSer 271
DB 181 CAATGGATGATAAAGACAGATGTAGGAGAGAGAGCCACTGGATCGCCCTCACAG 240
QY 272 GluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLys 291
DB 241 GAGCTGTAATGAATGGAAGTGGTGGATGGACATCTCCAGACTACAAAATTTG 300
QY 292 AlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeu 311
DB 301 GCTGGACAGCCGGAACTGGGGTCTATGCCCATGGCCAGGAGAGACTGTCTGGGTTG 360

QY 312 IleTyRAlaGlyClnTrpAsnAspPheGlnCysGluAspValAsnAsnPhe-IleCysG1 331
 |||||
 Db 361 ATTATGCTGGGAGGGAACGATTTCACATGTAAGACCTCAATAACTTAATTGGCA 420
 |||||
 QY 331 uLysAspArgGluThrValLeuSerSerAlaLeu 342
 |||||
 Db 421 AAAAGACAGGAGACAGTACTGCATCTGCATTA 454
 |||||

RESULT 10
 BB248064
 LOCUS
 DEFINITION BB248064 RIKEN full-length enriched, 7 days neonate cerebellum Mus
 musculus cDNA clone A730023E20 3', mRNA sequence.
 ACCESSION BB248064
 VERSION
 KEYWORDS
 SOURCE EST.
 ORGANISM house mouse.
 Mus musculus

REFERENCE
 AUTHORS 1 (bases 1 to 638)
 Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
 Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda,
 M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
 Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki,
 D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
 Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,
 Muramatsu, M. and Hayashizaki, Y.,
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
 Unpublished (2001)

TITLE
 JOURNAL
 COMMENT On Jul 6, 2000 this sequence version replaced gi:8940810.
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suhei-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Fax: 81-45-503-9222
 Tel: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura,
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara,
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa,
 K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
 Hayashizaki, Y.
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
 further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

FEATURES
 Location/Qualifiers
 1..638
 /organism="Mus musculus"

/db_xref="taxon:10090"
 /clone="A730023E20"
 /clone_lib="RIKEN full-length enriched, 7 days neonate
 cerebellum"
 /tissue_type="cerebellum"
 /dev_stage="7 days neonate"
 /lab_host="DH10B"
 /note="Site_1: SalI; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAAGGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 20.0 and subtraction to Rot = 459.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5' GAGAGAGATTCGAGTTAATAATTAATACCCCCCCCCC
 3']. cDNA was cleaved with XhoI and BamHI. Vector: a
 modified pBluescript KS(+) after bulk excision from Lambda
 FLC I."

BASE COUNT 212 a 134 c 146 g 146 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,69e-45 Length: 638
 Score: 831.00 Matches: 142
 Percent Similarity: 93.25% Conservatives: 10
 Best Local Similarity: 87.12% Mismatches: 11
 Query Match: 42.84% Indels: 0
 DB: 10 Gaps: 0

US-09-763-712a-2_COPY_206_547 (1-342) x BB248064 (1-638)

QY 180 ProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThr 199
 |||||
 Db 3 CTGCGCCCTCCAGGCGCCCTCAGGAGCAATGGAGCCATGGCTCTGCAGAATGAACCAACC 62
 |||||

QY 200 ProAlaProGluAspAsnGlyCysProHisTrpLysAsnPhetrAspLysCysTyr 219
 |||||
 Db 63 CCAGATCATGAGGTCACCGGATGTCGCTCCTGGAAGAACTTCACAGATAAATGCTAC 122
 |||||

QY 220 TyrPheSerValGluLysGluLeuPheGluAspAlaLysLeuPheCysGluAspLysSer 239
 |||||
 Db 123 TATTTTTCATTCGAAAAGAAATTTTGAAGATGCTAACTTTTCTGCGAAGACAATCT 182
 |||||

QY 240 SerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetVal 259
 |||||
 Db 183 TCCCATCTCGTTTTCATAAATCAAGAGAAGAACAGCAATGGATAAAGCATACCGTG 242
 |||||

QY 260 GlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrp 279
 |||||
 Db 243 GGGAGAGAAAGCCATTGGATCGGCTCAGAGACTCAGAAACAGGAAGCGAATGGAAGTGG 302
 |||||

QY 280 LeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyClnProAspAsnTrpGly 299
 |||||
 Db 303 CTAGACGGGTCACTCTGTGTTTACAAAACCTGGAAGCTTGGAAACACCACTAATCTGGGCG 362
 |||||

QY 300 HisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAsp 319
 |||||
 Db 363 AGTGCCCATGGCCAGGAGAACACTGTGCTGGCTTGATTTACGAGGACAGTGAATGAC 422
 |||||

QY 320 PheGlnCysGluAspValAsnAsnPhelIeCysGluLysAspArgGluThrValLeuSer 339
 |||||
 Db 423 TTCCAGTGTGTAATCAATAACTTCATTTGTGAGAGAGAAAGGAGGAGTACCATCA 482
 |||||

QY 340 SerAlaLeu 342
 |||||
 Db 483 TCCATATTA 491
 |||||

```

RESULT 11
BQ771366      808 bp      mRNA      linear      EST 26-JUL-2002
LOCUS        UI-M-F10-byu-g-09-0-UI.r1 NIH_BMAP_F10 Mus musculus CDNA clone
DEFINITION   IMAGE:5702432 5', mRNA sequence.
ACCESSION   BQ771366
VERSION     BQ771366.1 GI:21979842
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus.
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            NIH-MGC http://mgc.nci.nih.gov/.
            1 (bases 1 to 808)
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
Seq primer: pYX-5.
            Location/Qualifiers
            1..808
            /organism="Mus musculus"
            /strain="CS7BL/6"
            /db_xref="taxon:10090"
            /clone="IMAGE:5702432"
            /clone_lib="NIH_BMAP_F10"
            /tissue_type="whole brain"
            /dev_stage="embryo 12.5dpc"
            /note="Organ: Brain; Vector: pYX- Asc; Site.1: Ecor I;
            Site.2: Not I; The library was constructed according
            1996. Denatured RNA was size fractionated on a 1% agarose
            gel. First strand cDNA synthesis was primed with oligo-dT
            primer containing a Not I site. Double strand cDNA was
            size selected according to mRNA size fraction, ligated
            with Ecor I adaptor, digested with NotI and then cloned
            directionally into pYX-Asc vector. The library tag
            sequence located between the Not I site and the polyA tail
            is CAGCCAGCAC. This library was created for the University
            Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
            Developing Mouse Nervous System', supported by National
            Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
            program coordinator."
BASE COUNT   270 a 158 c 173 g 206 t 1 others
ORIGIN
Alignment Scores:
Pred. No.:      3,22e-41      Length:      808
Score:          767.00      Matches:    131
Percent Similarity: 93.33%      Conservat: 9
Best Local Similarity: 87.33%      Mismatches: 10
Query Match:    39.54%      Indels:    0
DB:            14      Gaps:      0
US-09-763-712a-2_COPY_206_547 (1-342) x BQ771366 (1-808)
QY 193 AlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProHisTrpLys 212
      |||||||
Db 2 GCTCTGCAGATGACCAACCCAGATGCAGAGTCCGCGCTCAGTGGAG 61
QY 213 AsnPheThrAspLysCysTrpTyrPheSerValGluLysGluIlePheGluAspAlaLys 232
      |||||||
Db 62 AACTTTCACAGATAAATGCTACTATTATTTTCATTTGGAAGAAATTTTGAAGATGCTAAG 121

```

```

QY 233 LeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArgGluGluGlnGln 252
      |||||||
Db 122 CTTTCTCTGAGACAAATCTTCCCATCTCGTTTTCATAAACTCAGAGAGAACAGCAA 181
      |||||||
QY 253 TrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGlu 272
      |||||||
Db 182 TGGATAAAAAAGCATACCGTGGGAGAGAAAGCCATTGGATCGGCCTCACAGACTCAGAA 241
      |||||||
QY 273 ArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAla 292
      |||||||
Db 242 CAGGAAGCGAATGGAAGTGGCTAGACGGGTCCCTGTGATTACAAAAACCTGGAAGCT 301
      |||||||
QY 293 GlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIle 312
      |||||||
Db 302 GGACAACCAAGATACTGGGCAGTGGCCATGGCCAGGAGAGACTGTCTGCTGTGATT 361
      |||||||
QY 313 TyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLys 332
      |||||||
Db 362 TACGAGGACAGTGGATGACTTCCAGTGTGAATCAATCAATCAATCAATCAATCAATCA 421
      |||||||
QY 333 AspArgGluThrValLeuSerSerAlaLeu 342
      |||||||
Db 422 GAAGGGAGGAGCAGTACCATCATCATATTA 451
      |||||||
RESULT 12
BQ891432      936 bp      mRNA      linear      EST 16-AUG-2002
LOCUS        AGENCOURT_8754195 NIH_MGC_130 Mus musculus CDNA clone IMAGE:6332754
DEFINITION   5', mRNA sequence.
ACCESSION   BQ891432
VERSION     BQ891432.1 GI:22283446
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            NIH-MGC http://mgc.nci.nih.gov/.
            1 (bases 1 to 936)
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
            Ph.D.
            CDNA Library Preparation: ResGen, Invitrogen Corp
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13790 row: n column: 19
            High quality sequence start: 25
            High quality sequence stop: 632.
FEATURES
            Location/Qualifiers
            1..936
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:6332754"
            /clone_lib="NIH_MGC_130"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: otocysts; Vector: pCMV-SPORT6.1.ccdB;
            Site.1: EcorV; Site.2: NotI; Cloned unidirectionally.
            Primer: Oligo dT. Average insert size 1.95 kb.
            Constructed by ResGen, Invitrogen Corp. Note: this is a
            NIH_MGC Library."
BASE COUNT   273 a 250 c 244 g 163 t 6 others
ORIGIN
Alignment Scores:
Pred. No.:      1.52e-36      Length:      936
Score:          696.50      Matches:    137
Percent Similarity: 91.72%      Conservat: 7

```


cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 The following repetitive elements were found in this cDNA
 sequence: 443-498, >POLY_A#Simple_repeat
 Seq primer: M13 Reverse.

```

/organism="Homo sapiens"
/db_xref="taxon.9606"
/clone="UI-E-EJ0-abq-c-05-0-UI"
/clone_lib="UI-E-EJ0"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EJ0 is a subcloned cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AAGATCAAGA
; lens, CCATTAGGCA; eye anterior segment, AATGCCGCAT;
optic nerve, CCATTAGGTC; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
184 a 89 c 121 q 112 t 1 others

```

| | |
|--|---|
| ORGANISM | Xenopus laevis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus. 1 (bases 1 to 591) |
| REFERENCE | NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . |
| AUTHORS | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), |
| TITLE | Tumor Gene Index |
| JOURNAL | Unpublished (1997) |
| COMMENT | Contact: Robert Strausberg, Ph.D. Email: cgapsb-r@mail.nih.gov Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov High quality sequence stop: 441. |
| FEATURES | Location/Qualifiers 1..591 /organism="Xenopus laevis" /db_xref="taxon:8355" /clone="IMAGE:4963626" /clone_lib="NICHG XGC Spl" /dev_stage="adult" /lab_host="DH10B (phage-resistant)" /note="Organ: spleen; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.4 kb. Constructed by Life Technologies." |
| BASE COUNT | 174 a 118 c 160 g 139 t |
| ORIGIN | |
| Alignment Scores: | |
| Pred. No.: | 3.38e-32 Length: 591 |
| Score: | 627.50 Matches: 112 |
| Percent Similarity: | 71.20% Conservative: 24 |
| Best Local Similarity: | 58.64% Mismatches: 46 |
| Query Match: | 32.35% Indels: 9 |
| DB: | 13 Gaps: 3 |
| US-09-763-712A-2_COPY_206_547 (1-342) x BI45884 (1-591) | |
| Qy 126 ProglyProProGlyProProGlyLyysGluGlyLeuProGlyProGlyProGly 145 | |

```

Db      25  CCAGTGGTCCCTGGGGCCCTGGAATAATGGAATCTTGGGCTTCTGTCGCCAGGA 84
QY      146  PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeu--- 164
Db      85  CCACAGGGGAAAGAGTCAATGGGAACTGGAGCTCCAGGGATTAGAGGGTCATCT 144
QY      165  -----ProGlyLeuProGlyValProGlyMetProGlyProGlyProGlyPro 182
Db      145  GGACCTCCAGGAATACCTGGATTACCTGGAAATACCTGGACCTAGAGTCTCCAGGGCCT 204
QY      183  ProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro----- 200
Db      205  CCTGGAGATTGGGAAGT-----TCTTTAGCATTACATCAGTCTCTTCAGCAGAAGTA 258
QY      201  -----AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCys 218
Db      259  ATGCTTGAACCTGTGGTGACCGTTCCTCTGGAGTGGGAACTTTCTGATAATGC 318
QY      219  TyrTyrPheSerValGluLysGluLeuPheGluAspAlaLysLeuPheCysGluAspLys 238
Db      319  TACTACTTCTCAACTGGAAGATATATTGATGATGCCAACTGATCTGTAAGAAAAG 378
QY      239  SerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLysGlnMet 258
Db      379  GCAGCGATGTTAGTTGTGATTGAACCCCAAGCAGCAGCAACAATTTTGAAGAAACAACA 438
QY      259  ValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLys 278
Db      439  AGTGGAAAGGAACCTTCTGGATAGTCTCAGAGATCTGAGGAGAGGGGTACATGGAAA 498
QY      279  TrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrp 298
Db      499  TGGCTGGATGGTACTACTCTAAACTACAAAAACTGGAAGAAAGGACAGCCGCGCGATTGG 558
QY      299  GlyHisGlyHisGlyProGlyGluAspCysAla 309
Db      559  TCTCATCACACTGGACCGAGGAAGATTGCGCT 591

```

Search completed: March 21, 2003, 12:20:33
Job time : 1136.43 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run On: March 21, 2003, 02:20:06 ; Search time 239.704 Seconds
(without alignments)
4293.470 Million cell updates/sec

Title: US-09-763-712A-2_COPY_91_547
Perfect score: 2520
Sequence: 1 MNSOLNSFTGOMENTTISQ.....EDVNFNICEKDRVTLSSAL 457

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/uspto_spool/US09763712/runat_14032003_100949_18091/app_query.fasta_1.1877
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=ptc -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712.ecgn.1.1.626.etrnat_14032003_100949_18091 -NCPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|--------|---------------|--------|-------|------------------------------|
| 1 | 2520 | 100.0 | 2024 | 21 | AAA07697 Human collectin en |
| 2 | 2520 | 100.0 | 2262 | 24 | ABA97932 Human scavenger re |
| 3 | 2520 | 100.0 | 2628 | 22 | AAH43036 Nucleotide sequenc |
| 4 | 2514 | 99.8 | 2181 | 22 | AAI60628 Human polynucleoti |
| 5 | 2514 | 99.8 | 2641 | 22 | AAC66903 Human EXMAD-14 cod |
| 6 | 2514 | 99.8 | 2929 | 24 | ABQ92072 Human polynucleoti |
| 7 | 2514 | 99.8 | 2930 | 20 | AAV55746 Human secreted pro |
| 8 | 2507 | 99.5 | 2318 | 22 | AAI58842 Human polynucleoti |
| 9 | 2487 | 98.7 | 2005 | 24 | ABL95574 Human angioogenesis |
| 10 | 2487 | 98.7 | 2005 | 24 | ABL88085 Human PRO7223 CDNA |
| 11 | 2344 | 93.0 | 2637 | 22 | AAH43037 Nucleotide sequenc |
| 12 | 1738 | 69.0 | 2256 | 22 | AAH43054 Nucleotide sequenc |
| 13 | 1715.5 | 68.1 | 1521 | 23 | AAH71133 DNA encoding novel |
| 14 | 682 | 27.1 | 3685 | 20 | AAH27858 Human CSR3 protein |
| 15 | 682 | 27.1 | 3810 | 20 | AAH27856 Human CSR1 protein |
| 16 | 474.5 | 18.8 | 1410 | 24 | ABK84622 Human CDNA differe |
| 17 | 474.5 | 18.8 | 1410 | 24 | ABL66964 Lung cancer relate |
| 18 | 471 | 18.7 | 5676 | 15 | AAQ64556 Human collagen (Ty |
| 19 | 471 | 18.7 | 8284 | 22 | AAH26526 Human breast cance |
| 20 | 471 | 18.7 | 8284 | 22 | AAH26553 Human breast cance |
| 21 | 471 | 18.7 | 8284 | 22 | AAH26600 Human breast cance |
| 22 | 462 | 18.3 | 4428 | 22 | AAD06574 Bovine alphas(III) |
| 23 | 462 | 18.3 | 4428 | 22 | AAD06575 Bovine alphas(III) |
| 24 | 459.5 | 18.2 | 756 | 14 | AAQ43034 Collagen-like poly |
| 25 | 459.5 | 18.2 | 756 | 17 | AAH16768 Collagen-like poly |
| 26 | 458.5 | 18.2 | 4821 | 23 | AAH86866 DNA encoding novel |
| 27 | 454 | 18.0 | 3171 | 21 | AAH12503 DNA encoding a hu |
| 28 | 454 | 18.0 | 3174 | 21 | AAH29843 DNA encoding huma |
| 29 | 453 | 18.0 | 6158 | 24 | ABL62095 Colon adenocarcino |
| 30 | 453 | 18.0 | 6158 | 24 | ABL65452 Lung cancer relate |
| 31 | 453 | 18.0 | 6158 | 24 | ABK33486 Human endometrial |
| 32 | 452.5 | 18.0 | 3170 | 21 | AAH12493 CDNA encoding huma |
| 33 | 452.5 | 18.0 | 3171 | 21 | AAH12502 CDNA encoding a hu |
| 34 | 452.5 | 18.0 | 3181 | 19 | AAV59358 Nucleotide sequenc |
| 35 | 452.5 | 18.0 | 3349 | 17 | AAT16518 Collagen A1/decori |
| 36 | 452.5 | 18.0 | 3349 | 21 | AAH12500 CDNA encoding a ch |
| 37 | 452.5 | 18.0 | 3531 | 21 | AAH12497 CDNA encoding a ch |
| 38 | 452.5 | 18.0 | 3535 | 17 | AAH16515 Collagen A1/BMP-2B |
| 39 | 452.5 | 18.0 | 3541 | 17 | AAH16516 Collagen A1/TGF-be |
| 40 | 452.5 | 18.0 | 3541 | 21 | AAH12498 CDNA encoding a ch |
| 41 | 452.5 | 18.0 | 4192 | 17 | AAT16517 Collagen A1/decori |
| 42 | 452.5 | 18.0 | 4409 | 19 | AAV60814 Human recombinant |
| 43 | 452.5 | 18.0 | 4770 | 22 | AAH22441 Human CDNA encodin |
| 44 | 452.5 | 18.0 | 6728 | 22 | AAF90491 Human pro-alpha-1 |
| 45 | 452.5 | 18.0 | 6728 | 24 | ABN97451 Gene #3949 used to |

ALIGNMENTS

RESULT 1
AAA07697
ID AAA07697 standard; DNA; 2024 BP.
XX
AC AAA07697;
XX
DT 20-JUN-2000 (first entry)
XX
DE Human collectin encoding DNA.
XX
KW Collectin; human; antibacterial; antiviral; ds.
OS Homo sapiens.
XX
FH Key
FT CDS
FT
FT Location/Qualifiers
55..1698
/*tag= a
/product= "collectin"

| | | | |
|--|---|---|------|
| QY | 381 | IlleClyLeuThrAspSerGluArgGluAenGluTrpLysTrpLeuAspGlyThrSerPro | 400 |
| Db | 1465 | ATCGGCTCACAGACTCAGAGCGTGAAAAATGAATGAAGTGGCTGGATGGACATCTCCA | 1524 |
| QY | 401 | AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly | 420 |
| Db | 1525 | GACTACAAAAATTGGAAGACTGGACAGCCGGATACTGGGGTCATGGCCATGGGCCAGGA | 1584 |
| QY | 421 | GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal | 440 |
| Db | 1585 | GAAGACTGTGCTGGTTGATTTATGCTGGGCAGTGAACGATTTCCTCAATGTGAAGACGTC | 1644 |
| QY | 441 | AsnAspPheIleCysGlnLysAspArgGluThrValLeuSerSerAlaLeu | 457 |
| Db | 1645 | AATAACTTTCATTTTGCAGAAAAAGACAGGGAGACAGTACTGTCTATCTGCATTA | 1695 |
| RESULT 2 | | | |
| ID | ABA97932 | standard; cDNA; 2262 BP. | |
| XX | ABA97932; | | |
| AC | ABA97932; | | |
| XX | | | |
| DT | 25-APR-2002 | (first entry) | |
| XX | | | |
| DE | Human scavenger receptor-like protein encoding cDNA SEQ ID NO 1. | | |
| XX | | | |
| KW | Human; scavenger; receptor; diagnosis; treatment; autoimmune disease; rheumatoid arthritis; gene; ss. | | |
| KW | | | |
| XX | | | |
| OS | Homo sapiens. | | |
| XX | | | |
| FF | Key | Location/Qualifiers | |
| FF | FT | 33..2262 | |
| FT | CDS | /*tag= a | |
| FT | | /product= "scavenger receptor-like protein" | |
| XX | | | |
| PN | JP2001340089-A. | | |
| XX | | | |
| PD | 11-DEC-2001. | | |
| XX | | | |
| PF | 08-DEC-2000; 2000JP-0375066. | | |
| XX | | | |
| PR | 27-MAR-2000; 2000JP-0090772. | | |
| XX | | | |
| PA | (SHIO) SHIONOGI & CO LTD. | | |
| XX | | | |
| DR | WPI; 2002-144965/19. | | |
| DR | P-PSDB; ABB08642. | | |
| XX | | | |
| PT | New scavenger receptor-like protein for diagnosis, prevention and treatment of autoimmune disease, such as rheumatoid arthritis - | | |
| PS | Claim 3; Fig 1; 38pp; Japanese. | | |
| XX | | | |
| CC | The invention relates to a human scavenger receptor-like protein. The protein is useful as a target molecule for diagnosis, prevention and treatment of autoimmune diseases such as rheumatoid arthritis. | | |
| XX | | | |
| SQ | Sequence 2262 BP; 670 A; 576 C; 581 G; 435 T; 0 other; | | |
| Alignment Scores: | | | |
| Pred. No.: | 1.9e-80 | Length: | 2262 |
| Score: | 2520.00 | Matches: | 457 |
| Percent Similarity: | 100.00% | Conservative: | 0 |
| Best Local Similarity: | 100.00% | Mismatches: | 0 |
| Query Match: | 100.00% | Indels: | 0 |
| DB: | 24 | Gaps: | 0 |
| US-09-763-712A-2_COPY_91_547 (1-457) x ABA97932 (1-2262) | | | |
| QY | 1 | MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln | 20 |
| Db | 889 | ATGAACAGCGAGCTCAACTTTCACAGTCAAGTGGAGAACATCACCACATCTCTCAA | 948 |

QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
|||||
Db 2029 ATCGGCCTCACAGACTCAGAGCGTGAATGAATGAATGAATGGAGTGGATGCTCCA 2088
QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyProGly 420
|||||
Db 2089 GACTACAAAATTTGAAAGCTGGACCGCGGATACTGGGGTCACTGGCCATGGCCAGGA 2148
QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
|||||
Db 2149 GAAGACTGTGCTGGGTGATTATGCTGGGCAGTGGACGATTTCATGTCAAGACGTC 2208
QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
|||||
Db 2209 AATAACTTCATTGCGAAAGACAGGGAGACAGTACTGTCATCTGCATTA 2259
RESULT 3
AAH43036
ID AAH43036 standard; DNA; 2628 BP.
XX
AC AAH43036;
DT 15-OCT-2001 (first entry)
DE Nucleotide sequence of a human scavenger receptor.
KW Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;
KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
KW low density lipoprotein; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 74..2302
ET /*tag= a
ET /*product= "scavenger receptor"
XX
XX
PN WO200159107-A1.
PD 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-JP00874.
PF
XX 14-FEB-2000; 2000JP-0035155.
PR 10-OCT-2000; 2000JP-0309068.
XX
PA (FUSO) FUSO PHARM IND LTD.
XX
PI Wakamiya N;
XX
DR WPI; 2001-497076/54.
DR P-PSDB; AAG63346.
XX
XX New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
PT useful for treatment and diagnosis of diseases associated with oxidized
PT low-density lipoprotein accumulation -
XX
PS Claim 2; Page 79-84; 118pp; Japanese.
XX
CC The present sequence encodes a human scavenger receptor, designated
CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
CC are useful in clarifying the functions of macrophages and basal
CC immunity. They are also useful in the treatment, prevention, diagnosis
CC and investigation of diseases such as arteriosclerosis, diabetic
CC complications, bacterial infection and restenosis following angioplasty,
CC which are associated with accumulation of oxidized low density
CC lipoprotein and the binding of advanced glycation end-products into
CC cells.
XX
SQ Sequence 2628 BP; 783 A; 661 C; 656 G; 528 T; 0 other;
Alignment Scores:
Pred. No.: 2.13e-80 Length: 2628

Score: 2520.00 Matches: 457
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0
US-09-763-712a-2_COPY_91_547 (1-457) x AAH43036 (1-2628)
QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
|||||
Db 929 ATGAACGCCAGCTCAACTCATTCACAGGTGAGTGGAGAACATCACCATCTCTCAA 988
QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
|||||
Db 989 GCCAACAGCAGCAGAACCTGAAGACCTGCAGAGACTTACACAAAGATGCAGAGAATAACA 1048
QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
|||||
Db 1049 GCCATCAAGTTCACCACTGGAGGAACGCTTCCAGCTCTTTGAGACGGATATTGTGAAC 1108
QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
|||||
Db 1109 ATCATTAGCAATATCAGTTTACACAGCCACCACTGCGGACGCTGACCAAGCAATCTAAAT 1168
QY 81 GluValArgThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
|||||
Db 1169 GAAGTCAGGACACATGTCACAGATACCTTACCAACACACAGATGATCTGACCTCTTG 1228
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
|||||
Db 1229 AATAATACCTGGCCCAACATCCGTTGGATTCTGTTCTCTCAGGATGCAACAAGATTG 1288
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
|||||
Db 1289 ATGAGTCGAGGTTAGACACTCAAGTAGCAACTTATCATGTCATTATGAAGAAATGAAG 1348
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
|||||
Db 1349 CTAGTAGACTCCCAAGCATGGTCAGCTCATCAAGAATTTTACAAATCTACAAGGTCACCG 1408
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGlyAsn 180
|||||
Db 1409 GCCCCAGGGGTCCAAGAGGTACAGAGATCCCAGGGACCCCTGCGCCCACTGGCAAC 1468
QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGlyGluArgGly 200
|||||
Db 1469 AAGGCACAGAAAGAGAGAGAGGGGAGCCCTGGACCACTGGCCCTGCGGGTGAGAGAGC 1528
QY 201 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
|||||
Db 1529 CCAATTGGACCAAGCTGGTCCCCCGGAGAGCGTGGCGCAAAAGGATCTAAAGGCTCCAG 1588
QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
|||||
Db 1589 GCCCCAAAGGCTCCCGTGTTCCTTGGGAAGCCCGGCCCTCAGGGCCCCCAGTGGGAC 1648
QY 241 ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProProGly 260
|||||
Db 1649 CCAGGCCCCCGGGCCCAACAGGCAAAAGAGGACTCCCCGGCCCTCAGGGCCCTCTCTGC 1708
QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
Db 1709 TTCCAGGACTTTCAGGGCACCTTGGGAGCCTGGGGTGCCTGGACCTCGGGGACTGCCA 1768
QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProProGlyPro 300
|||||
Db 1769 GGCCTGCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCCCGGCCCTCTGCGCCCA 1828
QY 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
|||||
Db 1829 TCAGGAGCGGTGTCCTCCCTGGCCCTGCAGATGAGCCAAACCCCGCACCGGAGACAAT 1888
QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340

Db 1889 GGCTGCCCGCCTCAGTGGAGAACTTCACAGACAAATGCTACTATTTTCAGTTGAGAAA 1948
 QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIle 360
 Db 1949 GAAATTTTGGAGATCCAAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTGTTCATA 2008
 QY 361 AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 Db 2009 AACACTAGAGAGAACAGCAATGGATTAATAAACACAGATGGTAGGAGAGAGAGCCACTGG 2068
 QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 Db 2069 ATCGGCTCACAGACTCAGAGCGTGGAATGAATGAAGTGGCTGGTGGACATCTCCA 2128
 QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyGlyProGly 420
 Db 2129 GACTCAAAATGGAAAGCTGGACAGCCGGATAACTGGGGTCATGGCCATGGGCCAGGA 2188
 QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 Db 2189 GAAGACTGTGCTGGTGTGATTTATGCTGGCAGTGGACGATTTCCAAATGTGAAGACGTC 2248
 QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 Db 2249 AATAACTTCAATTTGCGAAAAAGACAGAGGAGACAGTACTGTCTATCTGCATTA 2299
 RESULT 4
 AA160628
 ID AA160628 standard; cDNA; 2181 BP.
 AC AA160628;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 4617.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Dimanac RT;
 XX
 DR WPI; 2001-442253/47.
 XX P-PSDB; AAM41472.
 XX
 DR Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX

PS Claim 1; SEQ ID NO 4617; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA157798-AA161369) with nootropic, and
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activ/inhibit activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: the sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 2181 BP; 664 A; 539 C; 535 G; 443 T; 0 other;
 Alignment Scores:
 Pred. No.: 3,01e-80 Length: 2181
 Score: 2514.00 Matches: 456
 Percent Similarity: 99.78% Conservative: 0
 Best Local Similarity: 99.78% Mismatches: 1
 Query Match: 99.76% Indels: 0
 DB: 22 Gaps: 0
 US-09-763-712a-2_COPY_91_547 (1-457) x AA160628 (1-2181)
 QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrSerGln 20
 Db 317 ATGAACAGCCAGCTCAACTCACTTCAGGTACAGTGGAGACATCACCCTATCTCTCAA 376
 QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
 Db 377 GCCACAGCAGCAGCACTTGAAGACCTTGCAGGACTTACACAAAGATGCAGAGATAGAACA 436
 QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnPheGluThrAspIleValAsn 60
 Db 437 GCCATCAAGTTCAACCACTGGAGGAACTCCACGCTCTTTGAGACGATATGTGAAC 496
 QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
 Db 497 ATCATTTAGCAATATATAGTTACAGACCCACCACCTCGGGAGCGTACCAGCAATCTAAAT 556
 QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
 Db 557 GAAGTCAGGACCACTTGCACAGATACCTTACCACAAACACACAGATGATCTGACCTCTTG 616
 QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
 Db 617 AATAATACCTTGCCCAACATCCCTTTGGATTTCTCTCAGGATGCAACAGATTTG 676
 QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
 Db 677 ATGAGGTCGAGGTTAGACACTGAAGTAGCAACTTATCATGATTTATGGAAGAAATGAAG 736
 QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
 Db 737 CTAGTAGACTCCAAAGCATGCTCAGCTCATCAAGAAATTTTACATACTACAGGTCCACCG 796
 QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProGlyAsn 180
 Db 797 GGGCCAGGGGTTCCAAAGAGGTGACAGAGATCCAGGGACCCCTGGCCCACTGGCAAC 856
 QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGluArgGly 200
 Db 857 AAGGGACAGAAAG 916
 QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 Db 916 AAGGGACAGAAAG

```
Db 917 CCAATTGGACAGCTGGTCCCTCCCGGAGAGCGTGGCGGCAAGGATCTAAAGGCTCCAG 976
QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
    |||||
Db 977 GGCCCCAAAGGCTCCCGTGGTTCCCTCGGGAAGCCCGGCCCTCAGGCCCCAGTGGGAC 1036
QY 241 ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProProGly 260
    |||||
Db 1037 CCAGGCCCCCGGCGCCACAGGCAAGAGGACTCCCGGCCCTCAGGCCCCCTCTGGC 1096
QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyValProArgGlyLeuPro 280
    |||||
Db 1097 TTCAGGACTTCAGGCACCGTGGGAGCCCTGGGCTGCCTGGACTCGGGACTGCGCA 1156
QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPro 300
    |||||
Db 1157 GGCTTGCTGGGTACCAAGCATGCCAGGCCCAAGGCCCGCCCGCCCTCTGGCCCA 1216
QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
    |||||
Db 1217 TCAGGAGCGGTGGTCCCTGGCCCTGCAGATGAGCCCAACCCCGCAGGAGACAAT 1276
QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTrpTyrPheSerValGluLys 340
    |||||
Db 1277 AGCTGCCGCTCCTCGGAAGAACTTCACAGACAAATGCTACTATTTTCAGTTGAGAA 1336
QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
    |||||
Db 1337 GAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAGCTTTCACATCTTGTTTCATA 1396
QY 361 AsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
    |||||
Db 1397 AACACTAGAGAGAGCAAGCAATGGATAAAAAACAGATGGTAGGAGAGAGCCACTGG 1456
QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
    |||||
Db 1457 ATCGGCTCACAGACTCAGAGCTGAAATGAATGAAGTGGCTGGTGGGATCTCCA 1516
QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyProGly 420
    |||||
Db 1517 GACTACAAAAATTGGAAGCTGGACAGCCGATACTGGGTGATGCCCATGGCCAGGA 1576
QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
    |||||
Db 1577 GAAGACTGTGCTGGGTGATTTATGCTGGCGAGTGGAGCAATTTCCAATGTGAAGCGTC 1636
QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
    |||||
Db 1637 AATACTTTCATTTCCGAAAAAGACAGGGAGACAGTACTGTCTCATCTGCATTA 1687
RESULT 5
AAC66903
ID AAC66903 standard; cDNA; 2641 BP.
XX
AC AAC66903;
XX
DT 27-MAR-2001 (first entry)
XX
DE Human EXMAD-14 coding sequence SEQ ID NO: 39.
XX
KW Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
KW inflammation; reproductive disorder; cardiovascular disorder;
KW immune disorder; musculoskeletal disorder; developmental disorder;
KW gastrointestinal disorder; cell proliferation disorder; ss.
OS Homo sapiens.
XX
PN WO200068380-A2.
XX
PD 16-NOV-2000.
XX
PF 10-MAY-2000; 2000WO-US12811.
XX
```

```
PR 11-MAY-1999; 99US-0133643.
PR 23-AUG-1999; 99US-0150409.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Bandman O, Hillman JL, Tang YT, Lal P, Yue H, Baughn MR, Lu DAM;
PI Azimzai Y;
XX
DR WPI; 2001-007395/01.
DR P-PSDB; AAB27236.
XX
PT Isolated polynucleotide encoding extracellular matrix or
PT adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
PT preventing disorders associated with expression of EXMAD such as
PT proliferative, immune and genetic disorders -
XX
PS Claim 4; Page 121-122; 129pp; English.
XX
CC The present invention provides the protein and coding sequences for 25
CC novel extracellular matrix and adhesion-associated proteins (EXMADS).
CC These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
CC EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
CC EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
CC EXMAD-20, EXMAD-21, EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
CC useful in the prevention and treatment of cancers, cell proliferation,
CC cardiovascular, reproductive, immune, musculoskeletal, developmental and
CC gastrointestinal disorders and inflammation.
XX
SQ Sequence 2641 BP; 823 A; 628 C; 608 G; 582 T; 0 other;

Alignment Scores:
Pred. No.: 3.46e-80 Length: 2641
Score: 2514.00 Matches: 456
Percent Similarity: 99.78% Conservative: 0
Best Local Similarity: 99.78% Mismatches: 1
Query Match: 99.76% Indels: 0
DB: 22 Gaps: 0

US-09-763-712A-2_COPY_91_547 (1-457) x AAC66903 (1-2641)
QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
    |||||
Db 567 ATGAACAGCGCAGCTCAACTCATTCACAGGTTCAGATGGAGAACATCACCATCTCTCTCAA 626
QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
    |||||
Db 627 GCCAACGACGAGAACCTGAAAGACCTGCAGGACTTACACAAAGATGCAGAGATAGAACA 686
QY 41 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 60
    |||||
Db 687 GGCATCAAGTTCAACCAACTGGAGGAACGCTTCCAGCTCTTTGAGACGGATATTGTGAAC 746
QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
    |||||
Db 747 ATCATTAGCAATATCATGTTACACAGCCCACTCCGAGCGCTGACCAATCTAAAT 806
QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
    |||||
Db 807 GAAGTCAGGACCACTTGCACAGATACCTTTACCAACACACAGATGATCTGACCTCTTG 866
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
    |||||
Db 867 AATAATACCTCGCCCAACATCCGTTTGGATTCTGTTCTCTCAGGATGCAACAAGATTG 926
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetLys 140
    |||||
Db 927 ATGAGGTCAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGAAGAAATGAAG 986
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
    |||||
Db 987 CTAGTAGACTCCCAAGCATGGTCAGCTCATCAAGAAATTTTACAATACTACAGGTCCACCG 1046
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsn 180
```

| | | | |
|--------------------------------------|------------------|--|------|
| Db | 1047 | GGCCCCAGGGGTCACAGAGGTGACAGAGATCCACAGGACCCCTGCCCAACTGGCAAC | 1106 |
| Qy | 181 | LYSGLYGLNLSGLYGLULYSGLYGLUProGlyProGlyProGlyProAlaGlyGluAtrGly | 200 |
| Db | 1107 | AAGGGACAAAGAGAGAAAGGGGAGCCTGGACCACCTGCCCTCGGGGTGAGAGAGC | 1166 |
| Qy | 201 | ProIleGlyProAlaGlyProProGlyGluAtrGlyGlyLysGlySerLysGlySerGln | 220 |
| Db | 1167 | CCAATTTGGACCAAGCTGGTCCCCCGGAGAGCGTGCGGCAAGAGATCTAAAGGCTCCAG | 1226 |
| Qy | 221 | GlyProLysGlySerAtrGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp | 240 |
| Db | 1227 | GGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGCCCCAGTGGGGAC | 1286 |
| Qy | 241 | ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProProGly | 260 |
| Db | 1287 | CCAGGCCCCCGGGCCACCAGGCCAAAGAGGACTCCCGGCCCTCAGGCCCCCTCTGGC | 1346 |
| Qy | 261 | PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProAtrGlyLeuPro | 280 |
| Db | 1347 | TTCCAGGGGACTTCAGGGCACCCGTTGGGAGCGTGGGGTGCCGTGGACCTCGGGACTGCCA | 1406 |
| Qy | 281 | GlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPro | 300 |
| Db | 1407 | GGCTTGCCCTGGGGTACAGGCATGCGAGGCCCCCAAGGGCCCCCGGCCCTCTCTGGCCCCA | 1466 |
| Qy | 301 | SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAtrAspAsn | 320 |
| Db | 1467 | TCAGGAGCGGTGGTGCCTCGCCCTGCAGAAATGAGCAACCCCGGCCACCGGAGGACAAT | 1526 |
| Qy | 321 | GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys | 340 |
| Db | 1527 | AGCTGCCCGCCCTCACTGGGAAGAACTTCACAGACAAATGCTACTATTTTCAGTTGAGAAA | 1586 |
| Qy | 341 | GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle | 360 |
| Db | 1587 | GAATATTTTGAGGATGCAAAAGCTTTCTGTGAAGCAAGTCTTCACATCTTGTGTTTCATA | 1646 |
| Qy | 361 | AsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyAtrGluSerHisTrp | 380 |
| Db | 1647 | AACACTAGAGAGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGG | 1706 |
| Qy | 381 | IleGlyLeuThrAspSerGluAtrGluAsnGluTrpLysTrpLeuAspGlyThrSerPro | 400 |
| Db | 1707 | ATCGGCTCCACAGACTCAGAGCGGTAAAAATGAATGGAAGTGGCTGGATGGACATCTCCA | 1766 |
| Qy | 401 | AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly | 420 |
| Db | 1767 | GACTACAAAAATTGGAAGAGCTGGACAGCCGGATAACTGGGTCAATGGCCATGGGCCAGGA | 1826 |
| Qy | 421 | GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal | 440 |
| Db | 1827 | GAAGACTGCTGGGTGATTATGCTGGCAGTGGACGATTTCCAAATGTGAGAGAGCTC | 1886 |
| Qy | 441 | AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu | 457 |
| Db | 1887 | AATAAATTCAATTTGCAAAAACAGAGGAGACAGTACTGTCTATCTGCATTA | 1937 |
| RESULT 6 | | | |
| ABQ92072 | | | |
| ID ABQ92072 standard; cDNA; 2929 BP. | | | |
| XX | AC | ABQ92072; | |
| XX | AC | | |
| DT | 04-OCT-2002 | (first entry) | |
| XX | Human | polynucleotide SEQ ID NO 69. | |
| XX | Human; | cytostatic; antirheumatic; antiarthritic; vulnery; analgesic; | |
| KW | antinfammatory; | antibacterial; immunosuppressive; antiparkinsonian; | |
| KW | neuroprotective; | nootropic; osteopathic; haemostatic; vasotropic; | |
| KW | antituler; | fungicide; antidiabetic; antiasthmatic; antiallergic; | |

| | |
|----|--|
| KW | immunostimulant; antiparasitic; secreted protein; transmembrane protein; cytokine; cell proliferation; cell differentiation; autoimmune disease; |
| KW | stem cell; growth factor; nervous system disease; neuropathy; |
| KW | Alzheimer's disease; Parkinson's disease; Huntington's disease; |
| KW | osteoporosis; severe combined immunodeficiency; SCID; infection; |
| KW | multiple sclerosis; rheumatoid arthritis; gene therapy; gene; ss. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| XX | US2002065394-A1. |
| XX | |
| XX | 30-MAY-2002. |
| XX | |
| XX | 22-DEC-2000; 2000US-0745763. |
| PF | |
| XX | |
| XX | 18-MAR-1998; 98US-0040963. |
| XX | |
| XX | (JACO/) JACOBS K. |
| PA | (MCCO/) MCCOY J M. |
| PA | (LAVA/) LAVALLIE E R. |
| PA | (COLL/) COLLINS-RACIE L A. |
| PA | (EVAN/) EVANS C. |
| PA | (MERB/) MERBERG D. |
| PA | (TREA/) TREACY M. |
| PA | (SPAU/) SPAULDING V. |
| XX | |
| PI | Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C; |
| PI | Merberg D, Treacy M, Spaulding V; |
| XX | |
| XX | WPI: 2002-582343/62. |
| DR | P-PSDB: ABP61859. |
| DR | |
| XX | |
| PT | Novel secreted or transmembrane protein and polynucleotide encoding the |
| PT | protein, useful for diagnosis and treatment of neurological disorders, |
| PT | cancer, autoimmune diseases, bone disorders and lung or liver fibrosis |
| PT | _ |
| XX | |
| PS | Claim 250; Page 231-232; 284pp; English. |
| XX | |
| CC | The invention relates to human secreted or transmembrane protein (I), |
| CC | their fragments and is encoded by specific complementary deoxyribonucleic |
| CC | acid (cDNA) inserts (II), where the protein is substantially free from |
| CC | other mammalian proteins. (I) are useful for preventing, treating or |
| CC | ameliorating a medical condition, especially immunological treatment or |
| CC | prevention of tumours. (I) exhibits activity relating to angiogenesis, |
| CC | cytokine, cell proliferation, cell differentiation, antiinflammatory, |
| CC | stem cell growth factor activity and activin or inhibin-related |
| CC | activities. (I) can be used to manipulate stem cells in culture to give |
| CC | rise to neuroepithelial cells that can be used to augment or replace |
| CC | cells damaged by illness, autoimmune disease, accidental damage or |
| CC | genetic disorders. (I) induces the proliferation of neural cells and |
| CC | regeneration of nerve and brain tissue and is useful for the treatment of |
| CC | central and peripheral nervous system diseases and neuropathies, such as |
| CC | Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic |
| CC | lateral sclerosis. (I) is involved in chemotactic or chemokinetic |
| CC | activity, regulation of haematopoiesis and is useful for treating myeloid |
| CC | or lymphoid cell disorders, platelet disorders such as thrombocytopaenia |
| CC | and for regeneration of bone, cartilage, tendon, ligament and/or nerve |
| CC | tissue growth and in tissue repair, healing of burns, incisions, ulcers, |
| CC | for treating osteoporosis, osteoarthritis, bone degenerative disorders or |
| CC | periodontal disease. (I) is also useful for gut protection or |
| CC | regeneration and treatment of lung or liver fibrosis, reperfusion injury |
| CC | in various tissues, various immune deficiencies and disorders including, |
| CC | severe combined immunodeficiency (SCID), bacterial or fungal infections, |
| CC | autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis, |
| CC | diabetes mellitus, myasthenia gravis, allergic reactions and conditions, |
| CC | such as asthma or other respiratory problems. (II) is useful to express |
| CC | recombinant protein, as markers for tissues in which the corresponding |
| CC | protein is preferentially expressed and in gene therapy. The present |
| CC | sequence is that of a polynucleotide of the invention. |
| XX | |
| SQ | Sequence 2929 BP; 901 A; 699 C; 692 G; 637 T; 0 other; |

Sequence 2929 BP; 901 A; 699 C; 692 G; 637 T; 0 other;

Alignment Scores:

Pred. No.: 3,74e-80 Length: 2929
 Score: 2514.00 Matches: 456
 Percent Similarity: 99.78% Conservative: 0
 Best Local Similarity: 99.78% Mismatches: 1
 Query Match: 99.76% Indels: 0
 DB: 24 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x ABQ92072 (1-2929)

| | | | |
|----|------|--|------|
| QY | 1 | MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrIleSerGln | 20 |
| Db | 921 | ATCAACAGCCAGCTCAACTCAATTCACAGGTTCAGATGGAGAACATCACTATCTCTCAA | 980 |
| QY | 21 | AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr | 40 |
| Db | 981 | GCCACGAGCAGAACCTGAAAGACCTGCAGGACTTACACAAAGATCGAGAAATAGAACA | 1040 |
| QY | 41 | AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn | 60 |
| Db | 1041 | GCCATCAAGTTCAACCAACTGGAGGACGCTTCCAGCTCTTTGAGACGGATATTTGTAAC | 1100 |
| QY | 61 | IleIleSerAsnIleSerThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn | 80 |
| Db | 1101 | ATCATTAGCAATATCAGTTACACAGCCACCACTCGGACGCTGACCAAGCAATCTAAAT | 1160 |
| QY | 81 | GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu | 100 |
| Db | 1161 | GAAGTCAGGACCACTTGCACAGATACCTTACCAACACACAGATGATCTGACCTCTTGG | 1220 |
| QY | 101 | AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu | 120 |
| Db | 1221 | AATAATACCTGGCCCAACATCCGTTTGGATTCTGTTCTCTCAGGATGCAACAAGATTG | 1280 |
| QY | 121 | MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetLys | 140 |
| Db | 1281 | ATGAGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAAGAAATGAAG | 1340 |
| QY | 141 | LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro | 160 |
| Db | 1341 | CTAGTAGACTCCCAAGCATGTGACGTCATCAAGAAATTTACATACTACAAGTCCACCG | 1400 |
| QY | 161 | GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsn | 180 |
| Db | 1401 | GGCCCCAGGGGTCACAGAGGTGACAGAGGATCCAGGACCCCTCGCCCAACTGGCAAC | 1460 |
| QY | 181 | LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGly | 200 |
| Db | 1461 | AAGGACAGAAAGAGAGAGAGGGGGAGCCCTGGACCACTGGCCCTCGGGGTGAGAGGC | 1520 |
| QY | 201 | ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerGlySerGln | 220 |
| Db | 1521 | CCAATTGGACCACTGGTCCCTCCCGGAGAGCGGTGGCGCAAGGATCTAAAGGCTCCAG | 1580 |
| QY | 221 | GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp | 240 |
| Db | 1581 | GGCCCCAAAGGCTCCGCTGGTTCCCTTGGGAAGCCCGGCCCTCAGGGCCCTCTGGC | 1640 |
| QY | 241 | ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProProGly | 260 |
| Db | 1641 | CCAGGCCCCCGGGCCCAACAGCAAGAGGAGCTCCCGGCCCTCAGGGCCCTCTGGC | 1700 |
| QY | 261 | PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro | 280 |
| Db | 1701 | TTCAGGGACTTCAGGGCACCGTTGGGAGCCCTGGGTGCTGGACCTCGGGGACTGCCA | 1760 |
| QY | 281 | GlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPro | 300 |
| Db | 1761 | GGCTTCCTGGGTACAGGCATGCCAGGCCCAAGAGGCCCCCGGCCCTCTCTGGCCCA | 1820 |
| QY | 301 | SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn | 320 |
| Db | 1821 | TCAGGAGCGGTGGTGGCCCTTGGCCCTTGCAGATGACCAACCCCGGACCGGAGACAAT | 1880 |

| | | | |
|----|------|---|------|
| QY | 321 | GlyCysProHisTrpLysAsnPheThrAspLysCysTrpTyrPheSerValGluLys | 340 |
| Db | 1881 | AGCTGCCCGCTCACTGGAGAACTTTCACAGACAAATGCTACTATTTTCAGTTGAGAAA | 1940 |
| QY | 341 | GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle | 360 |
| Db | 1941 | GAAATTTTGGAGATGCAAGCTTTCTGTGGAAGACAAAGTCTTCACATCTGTGTTTCATA | 2000 |
| QY | 361 | AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp | 380 |
| Db | 2001 | AACACTAGAGAGAAACAGCAATGGATAAAAAACAGATGTAGGGAGAGAGACCCACTGG | 2060 |
| QY | 381 | IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLysLeuAspGlyThrSerPro | 400 |
| Db | 2061 | ATCGGCTCACAGACTCAGAGCGTGAAATGAATGAAGTGGCTGGTGGGACATCTCCA | 2120 |
| QY | 401 | AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyGlyProGly | 420 |
| Db | 2121 | GACTACAAAAATTTGAAAGCTGGACACCGGATAACTGGGGTCATGGCCATGGCCAGGA | 2180 |
| QY | 421 | GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal | 440 |
| Db | 2181 | GAAGACTGTGCTGGGTGATTATGCTGGCAGTGGACGATTCACATGTGAAGAGCTC | 2240 |
| QY | 441 | AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu | 457 |
| Db | 2241 | AATAACTTCATTGGCCAAAAAGACAGGACAGACTGTGTCATCTGCATTA | 2291 |

RESULT 7

| | |
|----------|---|
| AAV55746 | |
| ID | AAV55746 standard; cdNA; 2930 BP. |
| XX | AAV55746; |
| XX | 23-MAR-1999 (first entry) |
| XX | Human secreted protein clone bv227_1 coding sequence. |
| XX | Secreted protein; human; nutritional supplements; immune stimulant; |
| KW | immune suppressor; haematopoiesis regulator; tissue growth; chemotaxis; |
| KW | activin/inhibin; chemokinesis; haemostasis; thrombolysis; |
| KW | receptor/ligand activity; anti-inflammatory; tumour inhibitor; |
| KW | cadherin/tumour invasion suppressor; ds. |
| XX | Homo sapiens. |
| OS | |
| Key | Location/Qualifiers |
| FT | CDS |
| FT | 67..693 |
| FT | /*tag= a |
| XX | WO9855614-A2. |
| PN | |
| XX | 10-DEC-1998. |
| PD | |
| XX | 01-JUN-1998; 98WO-US11210. |
| PF | |
| XX | 29-MAY-1998; 98US-0087255. |
| PR | 04-JUN-1997; 97US-0868696. |
| PR | 04-JUN-1997; 97US-0868697. |
| PR | 04-JUN-1997; 97US-0868698. |
| PR | 04-JUN-1997; 97US-0868699. |
| PR | 04-JUN-1997; 97US-0868900. |
| PR | 04-JUN-1997; 97US-0869191. |
| PR | 04-JUN-1997; 97US-0869192. |
| PR | 04-JUN-1997; 97US-0869193. |
| PR | 04-JUN-1997; 97US-0869194. |

(GEMY) GENETICS INST INC.

| | |
|----|--|
| XX | Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER; |
| PI | McCoy JM, Racie LA, Spaulding V, Treacy M; |

XX WPI: 1999-059912/05.
 DR P-PSDB; AAW73628.
 XX New polynucleotides encoding secreted human proteins - derived from
 PT human foetal brain, adult brain, foetal kidney, adult ovary, adult
 PT retina, adult placenta or adult uterus cDNA libraries
 XX
 PS Claim 26: Page 87-88; 127pp; English.

XX This sequence encodes a human secreted protein of the invention.
 CC This DNA sequence was isolated from a human adult brain cDNA
 CC library, and was designated clone bv227-1. The DNAs and proteins
 CC are predicted to have biological activities which would make them
 CC suitable for treating, preventing or ameliorating medical conditions in
 CC humans and animals, although no supporting data is given. Suggested
 CC activities include nutritional sources or supplements, immune
 CC stimulating or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity. The DNAs
 CC are also stated to be useful for gene therapy. A host cell transfected
 CC with the DNA, or its subfragments and variants is useful for recombinant
 CC production of the human secreted protein clones.

XX SQ Sequence 2930 BP; 902 A; 699 C; 692 G; 637 T; 0 other;

Alignment Scores:
 Pred. No.: 3,74e-80 Length: 2930
 Score: 2514.00 Matches: 456
 Percent Similarity: 99.78% Conservative: 0
 Best Local Similarity: 99.78% Mismatches: 1
 Query Match: 99.76% Indels: 0
 DB: 20 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x AAV55746 (1-2930)

QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
 DB 921 ATGAACAGCCAGCTCAACTCATTACAGCTCAGATGGAGAACATCACCCTATCTCTCAA 980

QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
 DB 981 GCCAAGCAGCAGCAAGCTGAAAGACCTGCAGGACTTACACAAAGATGCAGAGAATAAGAACA 1040

QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
 DB 1041 GCCATCAAGTTCACCAACATGGAGGAACGCTTCCAGCTCTTTGAGACGGATATTGTGAAC 1100

QY 61 IleIleSerAsnIleSerThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
 DB 1101 ATCATTAGCAATATACATTACAGCCCAACCTGCGGACGCTGACCAAGCAATCTAAT 1160

QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
 DB 1161 GAAGTCAGGACCACTTGACAGATACCCCTACCAACACACAGATCATCTGACCTCCTTG 1220

QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnAspLeu 120
 DB 1221 AATAATACCTGGCCAAACATCGTTGGATTCTGTTCTCTCAGGATGCAACAAAGATTGG 1280

QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
 DB 1281 ATGAGGTCAGGTTAGACACTGAGTAGCAACTTATCATGCTATTATGGAAGAAATGAAG 1340

QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
 DB 1341 CTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGAATTTTACAATACTACAAGGTCCACCG 1400

QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsn 180
 DB 1401 GGCCCCAGGGGTCCAAAGAGGTACAGAGGATCCCAAGGGACCCCTGGCCCCAACTGGCAAC 1460

QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGly 200
 DB 1461 AAGGGACAAAGAGAGAGAGGGGGAGCTGGACACCTGGCCCTCGGGGTGAGAGAGGC 1520

QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 DB 1521 CCAATTGGACACAGCTGGTCCCGCGGAGAGCGTGGCGCAAGGATCTAAAGGCTCCAG 1580

QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
 DB 1581 GCGCCCAAGGCTCCCGTGGTCCCTGGGAAGCCCGGCCCTCAGGGCCCCAGTGGGAC 1640

QY 241 ProGlyProProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProProGly 260
 DB 1641 CCAGGCCCGCGGCCACAGGCAAGAGGAGCTCCCGCGGCCCTCAGGGCCCCCTCTGCG 1700

QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 DB 1701 TTCCAGGGGACTTCAGGGGACCGTGGGGAGCGCTGGGTCCTGGACCTCGGGGACTGCCA 1760

QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
 DB 1761 GGTTCCTGGGTACAGGATGCCAGGCCCAAGGGCCCCCGGCCCTCTCTGCCCCA 1820

QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 DB 1821 TCAGGAGCGGTGGTGCCTCGCCCTGCAGAAATGAGCCCAACCCGCGCAGGAGACAAT 1880

QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTrpPheSerValGluLys 340
 DB 1881 AGCTGCCCGCTCCTCTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1940

QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
 DB 1941 GAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTCATTA 2000

QY 361 AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 DB 2001 AACACTAGAGAGAACAGCAATGGATAAANAACACAGATGGTAGGAGAGAGAGCCACTGG 2060

QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 DB 2061 ATCGGCTCACAGACTCAGAGCGTGAAATGAATGAAGTGGCTGGATGGGACATCTCCA 2120

QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyGlyProGly 420
 DB 2121 GACTCAAAAATTGGAAGCTGGACAGCCGGATAACTGGGGTCATGGCCATGGGCCAGGA 2180

QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 DB 2181 GAAGACTGTCTGGGTTGATTTATGCTGGCAGTGAACGATTTCCAAATGTGAAGAGCTC 2240

QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 DB 2241 AATAACTTCAATTTGCAAAAAGACAGGAGAGACAGTACTGTCTCATCTGCATTA 2291

RESULT 8

AAI58842

ID AAI58842 standard; cDNA; 2318 BP.

XX

AC AAI58842;

XX

DT 22-OCT-2001 (first entry)

XX

DE Human polynucleotide SEQ ID NO 1045.

XX

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia; ss.

XX Homo sapiens.
 XX WO200153312-A1.
 XX 26-JUL-2001.
 PD 26-DEC-2000; 2000WO-US34263.
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR P-PSDB; AAM39886.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Claim 1; SEQ ID NO 1045; 10078pp; English.
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX SQ Sequence 2318 BP; 706 A; 560 C; 548 G; 504 T; 0 other;
 Alignment Scores:
 Pred. No.: 5,52e-80 Length: 2318
 Score: 2507.00 Matches: 457
 Percent Similarity: 99.78% Conservative: 0
 Best Local Similarity: 99.78% Mismatches: 0
 Query Match: 99.48% Indels: 1
 DB: 22 Gaps: 0
 US-09-763-712A-2_COPY_91_547 (1-457) x AAI58842 (1-2318)
 QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
 DB ATGAACAGCGCAGCTCAACTCAATTCACAGGTCAGAGTGGAGAACATCACCACCTATCTCTCAA 376
 QY 21 -AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgTh 40
 DB GCCCAACGAGCAGAGACCTGAAAGACCTGCAGGACCTTACACAAAGATGCAGAGATAGAAC 436
 QY 40 rAlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAs 60
 DB ACCCATCAAGTTCAACCACTGGAGGACGCTTCCAGCTCTTTGAGCGGATATTGTGAA 496

QY 60 nIleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAs 80
 DB CATCATTAGCAATATCAGTTACACAGCCACACCTGCGAGCGCTCACCAGCAATCTAAA 556
 QY 80 nGluValArgThrThrCysThrAspThrLysHisThrAspAspLeuThrSerLe 100
 DB TGAAGTCAGAGCACCTTGACAGATACCTTACCAAAACACACAGATGATGTGACCTCCTT 616
 QY 100 uAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLe 120
 DB GAATAATACCTTGGCCCAACATCGTTTGGATCTGTTCTCTCAGGATGCAACAAGATTT 676
 QY 120 uMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetIly 140
 DB GATGAGTTCAGGTTAGACACTGAAGTAGCAACTTATCAGTGCATTTATGAGAAATGAA 736
 QY 140 sLeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPr 160
 DB GCTAGTAGACTCCAAGCATGGTCAGCTCATCAAGAAATTTTACAATATCTACAAGGTCACC 796
 QY 160 oGlyProArgGlyProArgLysArgGlySerGlnGlyProProGlyProThrGlyAs 180
 DB GGGCCCCAGGGTCCCAAGAGGTGACAGAGATCCCAGGAGCCCCCTGGCCCACTGGCAA 856
 QY 180 nLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgG 200
 DB CAAGGACACAAAGGAGAGAGGAGGAGCTGGACCATCTGGCCCTGCGGGTGAGAGAG 916
 QY 200 yProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerG 220
 DB CCCAATTTGACCATGCTGGTCCCGCCGAGAGCGTGGCGGCAAGGATCTAAAGGCTCCCA 976
 QY 220 nGlyProLysGlySerArgLysSerProGlyLysProGlyProGlnGlyProSerGlyAs 240
 DB GGGCCCCAAAGGCTCCCGTGTCCCTCGGAAGCCCGCCCTCAGGCGCCCTGAGTGGGA 1036
 QY 240 pProGlyProGlyProGlyProGlyLysGlyGluGlyLeuProGlyProGlnGlyProG 260
 DB CCAGGCCCCCGGGCCCCACAGGCAAGAGGAGTCCCGCGGCCCTCAGGGCCCTCTGTG 1096
 QY 260 yPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPr 280
 DB CTTCCAGGAGACTTCAGGGCACCCTTGGGGAGCGCTGGGTGCTTGGACCTCGGGGAGTGC 1156
 QY 280 oGlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPr 300
 DB AGGCTTGCCTGGGTACCAGGATGCCAGGCCCAAGGCCCGCCCGCGCCCTCTCTGGCCC 1216
 QY 300 oSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAs 320
 DB ATCAGGAGCGGTGGTCCCTCGCCCTGCAGAAATGAGCCCAAGCCCGGACCCGAGGACAA 1276
 QY 320 nGlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLy 340
 DB TGGCTGCCCGCCTCACTTGAAGAATCTCAGACAAATGCTACTATTTTTCAGTGAGAA 1336
 QY 340 sGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheI 360
 DB AGAAATTTTGAAGTGCAGAAAGCTTTCTGTGAAGACAAAGTCTTCACATCTTTGTTTCAT 1396
 QY 360 eAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTr 380
 DB AAACACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGAGGACCATG 1456
 QY 380 pIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPr 400
 DB GATCGCCCTCACAGACTCAGAGCGTGAAGCGTGAAGAAAGTGGTAGGAGGACATCTCC 1516
 QY 400 oAspTrpLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGl 420
 DB AGACTACAAAAATTTGAAAGCTGGACGCCGATTAACCTGGGTGCATGGCCATGGGCCAGG 1576
 QY 420 yGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVa 440

|||||
 Db 1577 AGAAGACTGTGGTGGTATTATGCTGGGAGTGAACGATTTCCATGTGAAGACGT 1636
 Qy 440 lAsnAsnPhelIcysGluLysAspArgGluThrValLeuSerAlaLeu 457
 |||||
 Db 1637 CAATACCTTCATTTCGGAAGAGAGAGGAGACAGTACTGTCTATCTGCATTA 1688
 RESULT 9
 ABL95574
 ID ABL95574 standard; cDNA: 2005 BP.
 XX
 AC ABL95574;
 XX
 DT 19-JUL-2002 (first entry)
 XX
 DE Human angiogenesis related cDNA PRO7223 SEQ ID NO: 27.
 XX
 KW Human; angiogenesis: PRO protein; cardiovascularisation; wound; cancer;
 KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
 KW caroliant; cytostatic; antiangiogenic; hypotensive; vulnerary;
 KW antiarteriosclerotic; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200208284-A2.
 XX
 PD 31-JAN-2002.
 XX
 PF 09-JUL-2001; 2001WO-US21735.
 XX
 PR 20-JUL-2000; 2000US-219556P.
 PR 25-JUL-2000; 2000US-220624P.
 PR 25-JUL-2000; 2000US-220664P.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 02-AUG-2000; 2000US-223695P.
 PR 17-AUG-2000; 2000US-0643657.
 PR 23-AUG-2000; 2000WO-US23522.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 07-SEP-2000; 2000US-230978P.
 PR 15-SEP-2000; 2000US-000000P.
 PR 18-SEP-2000; 2000US-0664610.
 PR 18-SEP-2000; 2000US-0665350.
 PR 24-OCT-2000; 2000US-242922P.
 PR 08-NOV-2000; 2000US-0709238.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 10-NOV-2000; 2000WO-US30873.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000US-0747259.
 PR 20-DEC-2000; 2000WO-US34956.
 PR 22-JAN-2001; 2001US-0767609.
 PR 28-FEB-2001; 2001US-0796498.
 PR 28-FEB-2001; 2001WO-US06520.
 PR 01-MAR-2001; 2001WO-US06666.
 PR 09-MAR-2001; 2001US-0802708.
 PR 14-MAR-2001; 2001US-0808689.
 PR 22-MAR-2001; 2001US-0816744.
 PR 05-APR-2001; 2001US-0828366.
 PR 10-MAY-2001; 2001US-0854208.
 PR 10-MAY-2001; 2001US-0854280.
 PR 25-MAY-2001; 2001US-0866028.
 PR 25-MAY-2001; 2001US-0866034.
 PR 25-MAY-2001; 2001WO-US17092.
 PR 30-MAY-2001; 2001US-0870574.
 PR 30-MAY-2001; 2001WO-US17443.
 PR 01-JUN-2001; 2001WO-US17800.
 PR 20-JUN-2001; 2001WO-US19692.
 PR 28-JUN-2001; 2001WO-US00000.
 XX
 PA (GETH) GENENTECH INC.
 PA (BAKE/) BAKER K P.
 PA (FERR/) FERRARA N.
 PA (GERB/) GERBER H.
 PA (GERR/) GERRITSEN M E.

(GODD/) GODDARD A.
 PA (GODO/) GODOWSKI P J.
 PA (GURN/) GURNEY A L.
 PA (HILL/) HILLAN K J.
 PA (MARS/) MARSTERS S A.
 PA (PANJ/) PAN J.
 PA (PAON/) PAONI N F.
 PA (STEP/) STEPHAN J F.
 PA (WATA/) WATANABE C K.
 PA (WILL/) WILLIAMS P M.
 PA (WOOD/) WOOD W I.
 XX
 PI Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
 PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
 PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
 XX
 DR WPI: 2002-171999/22.
 DR P-PSDB: ABB95436.
 XX
 PT One hundred and eighty seven nucleic acids encoding PRO polypeptides,
 PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
 PT infarction), endothelial or angiogenic disorders in a mammal -
 XX
 XX Claim 1; Fig 27; 567pp; English.
 XX
 PS The present invention provides the protein and coding sequences of human
 CC PRO proteins. These are useful for treating or diagnosing a
 CC cardiovascular, endothelial or angiogenic disorder, including cardiac
 CC hypertrophy, trauma, cancer, age-related macular degeneration,
 CC atherosclerosis, hypertension, arterial stenosis, rheumatoid arthritis,
 CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
 CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound
 XX healing. The present sequence is a coding sequence of the invention.
 SQ Sequence 2005 BP; 606 A; 493 C; 491 G; 415 T; 0 other;
 Alignment Scores:
 Pred. No.: 2.48e-79 Length: 2005
 Score: 2487.00 Matches: 455
 Percent Similarity: 99.56% Conservative: 0
 Best Local Similarity: 99.56% Mismatches: 2
 Query Match: 98.69% Indels: 1
 DB: 24 Gaps: 0
 US-09-763-712a-2_COPY_91_547 (1-457) x ABL95574 (1-2005)
 Qy 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
 Db 134 ATGAACAGCCAGCTCAATTCAGGTTCAGATGGAGAACATCACCACTATCTCTCAA 193
 Qy 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
 Db 194 GCCAAGCAGCAGACCTGAAACACCTGCAGGACTTACACAAAGATGCAGAGATAGAACA 253
 Qy 41 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 60
 Db 254 GCATCAAGTTCACCACTGGAGGAACCTTCCAGCTCTTTGAGACGATATTTGGAAC 313
 Qy 61 IleIleSerAsnIleSerTyThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
 Db 314 ATCATTAGCAATATCAGTTACAGCCACCCACCTGCGGACGCTGACCACTCTAAAT 373
 Qy 81 GluValArgThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
 Db 374 GAAGTCAGGACCATTCGACAGATACCTT-ACCAACACACAGATGATCTGACCTCTTG 432
 Qy 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
 Db 433 AATAATACCTGGCCACACATCCGTTTGGATTCTGTTCTCTCAGGATCAACAAGATTG 492
 Qy 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetIys 140
 Db 493 ATGAGGTCGAGGTTAGACACTGAAGTAGCAACTTATCATGATTATGAGAAATGAAG 552

QY 141 LeuValAspSerLysHisGlyGlnLeuLeuLysAsnPheThrIleLeuGlnGlyProPro 160
 Db 553 CTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGATTTTACAACTACTACAAGGTCACCG 612
 QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGlyAsn 180
 Db 613 GGGCCCCAGGGTCCAAAGAGTACAGAGATCCCAAGGACCCCTGGCCCACTGGCAAC 672
 QY 181 LysGlyGlnLysGlyGlnLysGlyGluProGlyProGlyProAlaGlyGluArgGly 200
 Db 673 AAGGACAGAAAGAGAGAGAGGGGAGCTGGACCATCTGGCCCTGGGGTGAGAGGC 732
 QY 201 ProLeGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 Db 733 CCAATTGGACCATGGTCCCTCCCGGAGAGCGTGGCGCAAGGATCTAAAGGCTCCAG 792
 QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
 Db 793 GGGCCCCAAGGCTCCCGTGGTTCCCTGGGAAGCCCGCCCTCAGGGCCCACTGGGAC 852
 QY 241 ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProProGly 260
 Db 853 CCAGGCCCCCGGGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 912
 QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 Db 913 TTCCAGGAGTTCAGGCGCACCTGGGGAGCGCTGGGGTGGCTGGACCTCGGGGAGTGC 972
 QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPro 300
 Db 973 GGCCTGGCTGGGGTACAGGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1032
 QY 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 Db 1033 TCAGGAGCGGTGGTGGCTGGCTGGCTGGAGATGAGCAACCCCGGACCGGAGGACAT 1092
 QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
 Db 1093 AGCTGCGGCGCTCACTGGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAG 1152
 QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
 Db 1153 GAATTTTGGAGTGAAGAGCTTTCTGTGAAGACAAAGTCTTCACATCTTGTTCATA 1212
 QY 361 AsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 Db 1213 AACACTAGAGAGAACACCAATGGATGATAAAAAACAGATGGTAGGAGAGAGACCTGG 1272
 QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 Db 1273 ATCGGCTCACAGCTCAGAGCGTGAAATGAATGGAAGTGGCTGGATGGGACATCTCCA 1332
 QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
 Db 1333 GACTACAAAAATGGAAGCTGGACGCGGATTAAGTGGGTCATGGCCATGGGCGACGA 1392
 QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 Db 1393 GAAGACTGTGCTGGGTGATTTATCTGCTGGGAGTGGAGACGATTTCCATGTAAGACGTC 1452
 QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 Db 1453 AATAACTTCATTGGGAAAAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1503
 RESULT 10
 ID ABL88085
 AC ABL88085;
 DT 16-MAY-2002 (first entry)
 XX

Human PRO7223 cDNA sequence SEQ ID NO:27.

Human; angiogenesis; cardiant; cytostatic; antiangiogenic; hypotensive;
 vulnery; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
 gene therapy; cardiovascular disorder; endothelial disorder; cancer;
 angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
 age-related macular degeneration; arterial restenosis; angina;
 rheumatoid arthritis; myocardial infarction; thrombophlebitis;
 lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
 wound healing; chromosome mapping; gene mapping; gene; ss.

Homo sapiens.

WO200200690-A2.

03-JAN-2002.

20-JUN-2001; 2001WO-US19692.

23-JUN-2000; 2000US-213637P.

20-JUL-2000; 2000US-219556P.

25-JUL-2000; 2000US-220624P.

25-JUL-2000; 2000US-220664P.

28-JUL-2000; 2000WO-US20710.

02-AUG-2000; 2000US-222695P.

17-AUG-2000; 2000US-0643657.

23-AUG-2000; 2000WO-US23328.

07-SEP-2000; 2000US-230978P.

18-SEP-2000; 2000US-0664610.

18-SEP-2000; 2000US-0665350.

24-OCT-2000; 2000US-242922P.

08-NOV-2000; 2000US-0709238.

10-NOV-2000; 2000WO-US30873.

01-DEC-2000; 2000WO-US32678.

20-DEC-2000; 2000US-0747259.

22-DEC-2000; 2000WO-US34956.

22-JAN-2001; 2001US-0767609.

28-FEB-2001; 2001US-0796498.

28-FEB-2001; 2001WO-US06520.

01-MAR-2001; 2001WO-US06666.

09-MAR-2001; 2001US-0802706.

14-MAR-2001; 2001US-0808689.

22-MAR-2001; 2001US-0816744.

05-APR-2001; 2001US-0828366.

10-MAY-2001; 2001US-0854208.

25-MAY-2001; 2001US-0866028.

25-MAY-2001; 2001US-0866034.

25-MAY-2001; 2001WO-US17092.

30-MAY-2001; 2001US-0870574.

30-MAY-2001; 2001WO-US17443.

01-JUN-2001; 2001WO-US17800.

(GETH) GENENTECH INC.

Baker KP, Ferrara N, Gerber H, Gerecht ME, Goddard A;

Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;

Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;

WPI; 2002-090516/12.

P-PSDB; ABB84830.

One hundred and eighty seven nucleic acids encoding PRO polypeptides,
 useful in diagnosis and treatment of cardiovascular (e.g. myocardial
 infarction), endothelial or angiogenic disorders in a mammal -

Claim 2; Fig 27; 565pp; English.

ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to

ABB85003. The PRO proteins and polynucleotides have cardiant, cytostatic,

antiangiogenic, hypotensive, vulnery and antiarteriosclerotic

CC activities, and can be used in gene therapy. The PRO polynucleotides,
 CC proteins, agonists and antagonists are useful for treating or diagnosing
 CC a cardiovascular, endothelial or angiogenic disorder in a mammal,
 CC e.g. cardiac hypertrophy, trauma, cancer, age-related macular
 CC degeneration, atherosclerosis, hypertension, arterial restenosis,
 CC rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis,
 CC lymphangitis, tumour angiogenesis (such as breast carcinoma and liver
 CC carcinoma) and wound healing. The PRO polynucleotides have applications
 CC in molecular biology, including use as hybridisation probes, and in
 CC chromosome and gene mapping. ABL88259 to ABL88267 represent primers and
 CC probes used in the exemplification of the present invention.

XX
 SQ Sequence 2005 BP: 606 A; 493 C; 491 G; 415 T; 0 other;

Alignment Scores:

Pred. No.: 2,48e-79 Length: 2005
 Score: 2487.00 Matches: 455
 Percent Similarity: 99.56% Conservative: 0
 Best Local Similarity: 99.56% Mismatches: 2
 Query Match: 98.69% Indels: 1
 DB: 24 Gaps: 0

US-09-763-712A-2_COPY_91_547 (1-457) x ABL88085 (1-2005)

QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
 DB 134 ATGAACAGCCAGCTCAACTCATTACAGTCCAGTGGAGAACATCACCACCTATCTCTCAA 193
 QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
 DB 194 GCCAACGAGCAAGCCTGAAGACCTGCGAGGACTTACACAAAGATGCGAGAGAAATAGACA 253
 QY 41 AlaIleLysPheAsnGlnLeuGluGlnArgPheGlnLeuPheGluThrAspIleValAsn 60
 DB 254 GCCATCAAGTTCAACCACTGGAGCAAGCGTCCAGCTCTTGAGCGGATATTGTGAAC 313
 QY 61 IleIleSerAsnIleSerTyrThrAlaHisLysLeuArgThrLeuThrSerAsnLeuAsn 80
 DB 314 ATCAATTAGCAATATCAGTTACAGCCACCCACCTGCGGACGCTGACCCAGCAATCTAAAT 373
 QY 81 GluValArgThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
 DB 374 GAAGTCAGACCACTTGACAGATACCTT-ACCAACACACAGATGATCTGACCTCCCTTG 432
 QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
 DB 433 AATAATACCTGGCCCAACATCCGTTGGATTCTGTTCTCTCAGGATGCAACAAAGATTG 492
 QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
 DB 493 ATGAGTTCGAGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAAGAAATGAAG 552
 QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
 DB 553 CTAGTAGACTCCAGCATGCTGAGTCATCAAGAATTTTACAATACTACAAGGTCCACCG 612
 QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsn 180
 DB 613 GSCCCAGGGTCCAAAGAGGTGACAGAGATCCAGGAGCCCTGGCCCACTGGCAAC 672
 QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAlaGlyGluArgGly 200
 DB 673 AAGGACAGAAAGAGAGAAAGGGGAGCGCTGGACCACTGGCCCTGCGGGTGAGAGAGGC 732
 QY 201 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 DB 733 CCATTTGACCAAGCTGGTCCCGGAGAGCGTGGCGCAAGGATCTAAAGGCTCCCGAG 792
 QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
 DB 793 GSCCCCAAGAGTCCCGTGGTTCCCTGGGAAGCGCCGCTCAGGGCCCACTGGGAC 852
 QY 241 ProGlyProProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProGly 260

DB 853 CCAGGCCCCCGGGCCACCAGGCAAGAGGAGCTCCCGGGCCCTCAGGGCCCTCTCTGC 912
 QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 DB 913 TTCCAGGAGACTTCAGGGACCCGTTGGGAGCCTGGGGTGCCTGGAGCTCGGGACTGCCA 972
 QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
 DB 973 GGCTTGGCTGGGTACAGGCATGCAGGCCCAAGGGCCCCCGGGCCCTCTGGCCCA 1032
 QY 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 DB 1033 TCAGGAGCGGTGGTCCCTGCGCAATGAGCAATGAGCAACCCCGGACCGGAGACAAT 1092
 QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
 DB 1093 AGCTGCCGCCCTCACTGGGAAGAACCTCAGAGACAATGCTACTATTTTCAGTTGAGAAA 1152
 QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
 DB 1153 GAAATTTTGGAGATGCAAGCTTTCTGTGAAGACAAGCTTTCACATCTTGTTCATA 1212
 QY 361 AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 DB 1213 AACACTAGAGAGAACAGCAATGGATAAAAAAACAGATGCTAGGAGAGAGAGCCACTGG 1272
 QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 DB 1273 ATCGGCCCTCACAGACTCAGAGCGTGAATGAATGAAGTGGTGGATGGGACATCTCCA 1332
 QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
 DB 1333 GACTACAAAAATTGAAAGCTGGACAGCGGATAACTGGGGTCATGGCCATGGCCAGCA 1392
 QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 DB 1393 GAAGACTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1452
 QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 DB 1453 AATACTTCATTTGCAAAAAGACAGGGAGACAGTACTGTCTCATCTGCATTA 1503
 RESULT 11
 AAH43037
 ID AAH43037 standard; cDNA; 2637 BP.
 XX
 AC AAH43037;
 XX
 XX 15-OCT-2001 (first entry)
 XX
 DE Nucleotide sequence of a human scavenger receptor.
 KW Human; scavenger receptor; SRCL-P1; macrophage; basal immunity;
 KW arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
 KW low density lipoprotein; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 CDS 92..2320
 FT /*tag= a
 FT /product= "scavenger receptor"
 XX
 PN WO200159107-A1.
 XX
 PD 16-AUG-2001.
 XX
 XX 08-FEB-2001; 2001WO-JP00874.
 XX
 XX 14-FEB-2000; 2000JP-0035155.
 PR 10-OCT-2000; 2000JP-0309068.
 XX

/product= "scavenger receptor"

FT WO200159107-A1.
 PN 16-AUG-2001.
 PD 08-FEB-2001; 2001WO-JP00874.
 XX 14-FEB-2000; 2000JP-0035155.
 XX 10-OCT-2000; 2000JP-0309068.
 PR (FUSO) FUSO PHARM IND LTD.
 PA Wakamiya N;
 XX WPI; 2001-497076/54.
 PI P-PSDB; AAG63350.
 XX
 DR New Scavenger receptor proteins SRCL-P1 with collectin-like structure,
 XX useful for treatment and diagnosis of diseases associated with oxidized
 PT low-density lipoprotein accumulation -
 XX Claim 4; Page 105-109; 118pp; Japanese.
 PS
 XX The present sequence encodes a human scavenger receptor, designated
 CC SRCL-P1. The SRCL-P1 polypeptide has a collectin-like structure. They
 CC are useful in clarifying the functions of macrophages and basal
 CC immunity. They are also useful in the treatment, prevention, diagnosis
 CC and investigation of diseases such as arteriosclerosis, diabetic
 CC complications, bacterial infection and restenosis following angioplasty,
 CC which are associated with accumulation of oxidized low density
 CC lipoprotein and the binding of advanced glycation end-products into
 CC cells.
 XX
 SQ Sequence 2256 BP; 721 A; 524 C; 528 G; 483 T; 0 other;

Alignment Scores:
 Pred. No.: 4.19e-53 Length: 2256
 Score: 1738.00 Matches: 333
 Percent Similarity: 72.87% Conservative: 0
 Best Local Similarity: 72.87% Mismatches: 0
 Query Match: 68.97% Indels: 124
 DB: 22 Gaps: 1

US-09-763-712a-2_COPY_91_547 (1-457) x AAH43054 (1-2256)

QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
 DB 929 ATGAACAGCCAGCTCAACTCATTCACAGGTCAGATGGAGAACATCACCACCTATCTCTCAA 988
 QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
 DB 989 GCCAAGGAGCAGACCTGAAAGACCTGCAGGACTTACACAAGATGCAGAGATAGACA 1048
 QY 41 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 60
 DB 1049 GCATCAAGTTCAACCAACTGGAGGAACGCTTCCAGCTTTGAGACGGATATTGTGAAC 1108
 QY 61 IleIleSerAsnIleSerThrAlaHisLysLeuArgThrLeuThrSerAsnLeuAsn 80
 DB 1109 ATCATAGCAATATCAGTTACACAGCCACCCACCTGCGGACGCTGACCAATCTAAAT 1168
 QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspLeuThrSerLeu 100
 DB 1169 GAAGTCAGGACCACTTGACAGATACCTTACCAACACACAGATGATCTGACCTCCTTG 1228
 QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
 DB 1229 AATAATACCTGGCCCAACATCGTTTGGATTCTGTTCTCTCAGGATGCAACAAGATTG 1288
 QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetLys 140
 DB 1289 ATGAGGTGAGGTTAGACACTGAAGTAGCAACTTATCAGTATTATGGAAGAAATGAAG 1348

QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
 DB 1349 CTAGTAGACTCCAGCATGGTCAGCTCATCAAGAATTTTACATACTACAGGTCCACCG 1408
 QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGlyAsn 180
 DB 1409 GGGCCAGGGTCCAAAGAGGTGACAGAGATCCAGGGACCCCTGGCCCACTGGCAAC 1468
 QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGluArgGly 200
 DB 1469 AAGGACAGAAAGAGAGAGGGGAGCTGGACCACTGGCCCTGCG - - - - - 1516
 QY 201 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 DB 1516 - - - - - 1516
 QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
 DB 1516 - - - - - 1516
 QY 241 ProGlyProProGlyProGlyLysGlyLysGlyLeuProGlyProGlnGlyProProGly 260
 DB 1516 - - - - - 1516
 QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 DB 1516 - - - - - 1516
 QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPro 300
 DB 1516 - - - - - 1516
 QY 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 DB 1516 - - - - - 1516
 QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
 DB 1517 GGCTGCCCGCCCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1576
 QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
 DB 1577 GAAATTTTGGAGATGCAAGCTTTTCTGTGAAGACAACTTTCACATCTTGTTCATTA 1636
 QY 361 AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 DB 1637 AACACTAGAGAGAACAGCAATGGATAAAACACAGATGGTAGGAGAGAGCCACTGG 1696
 QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 DB 1697 ATCGGCTCACAGACTCAGAGCGTGAATGAATGAAGTGGCTGGATGGACATCTCCA 1756
 QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
 DB 1757 GACTACAAAATTTGGAAGTGGACAGCGGATAACTGGGGTCATGGCCATGGGCCAGGA 1816
 QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 DB 1817 GAAGACTGTGTGGGTGTGATTTATGCTGGCAGTGGAGAGATTTCCAATGTGAAGACGTC 1876
 QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 DB 1877 ATATCTTCATTTGGCAAAAACAGACGGGAGACAGTACTGTCTCATCTGCATTA 1927
 RESULT 13
 ID AAS71133
 XX AAS71133 standard; cDNA; 1521 BP.
 AC AAS71133;
 XX
 DT 13-FEB-2002 (first entry)
 XX

DE DNA encoding novel human diagnostic protein #6937.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG06946.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 1; SEQ ID NO 6937; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AA564197-AA594564 represent novel human
 CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1521 BP; 386 A; 420 C; 425 G; 290 T; 0 other;

Alignment Scores:

Pred. No.: 1,92e-52 Length: 1521
 Score: 1715.50 Matches: 298
 Percent Similarity: 94.43% Conservative: 7
 Best Local Similarity: 92.26% Mismatches: 7
 Query Match: 68.08% Indels: 11
 DB: 23 Gaps: 1

US-09-763-712a-2_COPY_91_547 (1-457) x AAS71133 (1-1521)

QY 132 LeuSerValIleMetGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLys 151

Db 130 CTACATGTCTCTATTATGCAATGAGAAATCTACCCACAGG----- 171

QY 152 AsnPheThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySer 171

Db 172 -----CAAGGTCCACCGGGCCCGAGGGTCCCAAGAGGTACAGAGATCC 216

QY 172 GlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGlnLysGlyGluProGly 191

15-.. * *

| | | | | |
|-----------|--|--|--|------|
| Db | 217 | | CAGGGACCCCTGGCCCAACTGGCAACAAGGGACACAAAGGAGAGAGGGGGAGCCTGGA | 276 |
| QY | 192 | | ProProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArg | 211 |
| Db | 277 | | CCACCTGGCCCTGCGGGTGAGAGAGGCCCAATTTGGACCACTGGTCCCCCGGAGACGT | 336 |
| QY | 212 | | GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys | 231 |
| Db | 337 | | GGCGCAAGGATCTAAAGGCTCCCAAGGGCCCAAGGCTCCCGTGGTTCCTCCCTGGGAAG | 396 |
| QY | 232 | | ProGlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGluGly | 251 |
| Db | 397 | | CCCGGCCCTCAGGGCTCCAGTGGGACCCAGGCCCGCCCGGCCACCAAGCAAGAGGA | 456 |
| QY | 252 | | LeuProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluPro | 271 |
| Db | 457 | | CTCCCCGGCCCTCAGGGCCCTCTGGCTTCCAGGGACTTCAGGGGACCGTTGGGAGCCCT | 516 |
| QY | 272 | | GlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyPro | 291 |
| Db | 517 | | GGGGTGCCTGGACCTCGGGGACTGCCAGCTTGCCTGGGGTACCAGGCATGCCAGGCC | 576 |
| QY | 292 | | LysGlyProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsn | 311 |
| Db | 577 | | AAGGGCCCGCCCGCCCTCTGGCCCATCAGGAGCGGTGGTCCCTGGCCCTGCAGAAAT | 636 |
| QY | 312 | | GluProThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAsp | 331 |
| Db | 637 | | GAGCCAAACCCAGCACCGGAGGACAATGGTGGCCGCCCTCCTCGGAGAACTTTCAGAC | 696 |
| QY | 332 | | LysCysTrpTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGlu | 351 |
| Db | 697 | | AAATGCTACTATTTTCACTTGAGAAAGAAATTTTGGAGTATGCAAGCTTTTCTGTGAA | 756 |
| QY | 352 | | AspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLys | 371 |
| Db | 757 | | GACAAGTCTTCATCTTTTTCATAAACACTAGAGAGGAGAACAGCAATGGATAAAAAA | 816 |
| QY | 372 | | GlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGlu | 391 |
| Db | 817 | | CAGATGGTAGGAGAGAGAGAGCCACTGGATCGGCTCACAGCTCAGAGCGTGAAATGAA | 876 |
| QY | 392 | | TrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAsp | 411 |
| Db | 877 | | TGSAAGTGGCTGGATGGGACATCTCCAGACTACAAAAATTGGAAGCTGCAGACCGGAT | 936 |
| QY | 412 | | AsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGln | 431 |
| Db | 937 | | AACTGGGGTCATGGCCATGGCCAGGAGAGAGCTGTGCTGGGTGTGATTTATGCTGGGCAG | 996 |
| QY | 432 | | TrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGluThr | 451 |
| Db | 997 | | TGGAAGCATTTCCATGTGAAGACGTCATACTTCATTTCGAAAAAGACAGGGAGACA | 1056 |
| QY | 452 | | ValLeuSer 454 | |
| Db | 1057 | | GTCTTTCAGC 1065 | |
| RESULT 14 | | | | |
| AAAX27858 | | | | |
| ID | AAAX27858 | | standard; DNA; 3685 BP. | |
| XX | AAAX27858; | | | |
| AC | AAAX27858; | | | |
| DT | 02-JUN-1999 | | (first entry) | |
| DE | Human CSR3 | | protein coding sequence. | |
| XX | Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage; | | | |
| KW | scavenger receptor protein; intracellular stress; arteriosclerosis; | | | |
| KW | diabetic circulatory obstruction; microbial infection; ss. | | | |
| XX | | | | |

OS Homo sapiens.
 XX WO9909159-A1.
 XX 25-FEB-1999.
 XX 12-AUG-1998; 98WO-JP03602.
 XX 30-JUL-1998; 98JP-0230121.
 PR 13-AUG-1997; 97JP-0233396.
 XX (NISR) JAPAN TOBACCO INC.
 XX Nakamura Y, Tokino T;
 PI WPI; 1999-181032/15.
 DR P-PSDB; AAY00994.
 XX Scavenger receptor proteins - for treatment and diagnosis of
 PT disorders involving cell stress
 PS Claim 6; Page 142-150; 175pp; Japanese.
 XX This sequence encodes the human cellular stress response 3 (CSR3) protein
 CC of the invention. The CSR proteins are macrophage scavenger receptor
 CC proteins. The CSR proteins can be used in the treatment, gene therapy
 CC and diagnosis of diseases in which intracellular stress is important,
 CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial
 CC infection. Expression of the proteins is induced in vivo in response to
 CC intracellular stress, and inhibits cell death as a result of such stress.
 XX Sequence 3685 BP; 901 A; 1142 C; 972 G; 669 T; 1 other;
 SQ

Alignment Scores:
 Pred. No.: 5,04e-16 Length: 3685
 Score: 682.00 Matches: 132
 Percent Similarity: 62.42% Conservative: 54
 Best Local Similarity: 44.30% Mismatches: 94
 Query Match: 27.06% Indels: 18
 DB: 20 Gaps: 2

US-09-763-712A-2_COPY_91_547 (1-457) x AAX27858 (1-3685)

QY 11 GlnMetGluAsnIleThrThrIleSerGlnAlaAsnGluGlnAsnLeuLysAspLeuGln 30
 Db 1078 CAGCTGGATAACATCTCGTCCCTCTCGGATGACCAAGAGAGAACATGATGATCTTCAG 1137
 QY 31 AspLeuHisLysAspAlaGluAsnArgThrAlaIleLysPheAsnGlnLeuGluArg 50
 Db 1138 TACCATACCCACTACGCCCAAGACCGACCTGTGGAGAGGTTTGAGTCTCTGGAAGAGCGC 1197
 QY 51 PheGlnLeuPheGluThrAspIleValAsnIleIleSerAsnIleSerTyrThrAlaHis 70
 Db 1198 ATGGCTTCTCAGCAGATGAAATGGCCACCATCTTCACCAACATCAATGCCACCAACAC 1257
 QY 71 HisLeuArgThrLeuThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeu 90
 Db 1258 CACGTGCACAGCAGTCTCAAGTACCTGGATGACGTGCGCTCTCTGCAGCGTGGGCTTC 1317
 QY 91 ThrLysHisThrAspAspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAsp 110
 Db 1318 CACACCCATGCGGAGGAGCTTACTACCTGAACAAGTCTGTCTCCATCATGCTGGGCACC 1377
 QY 111 SerValSerLeuArgMetGlnAspLeuMetArgSerArgLeuAspThrGluValAla 130
 Db 1378 ACAGACCTGCTCCGGAGGCGTTACGCTGCTCACTGCCGCGCTGACCTCAACGTCGGG 1437
 QY 131 AsnLeuSerValIleMetGluGluMetLysLeuValAspSerLysHisGlnLeuIle 150
 Db 1438 AACCTCTCATGATCGTGGAGAGATGAAGGAGTGGACACACACATGGAGAAATCCTT 1497
 QY 151 LysAsnPheThrIleLeuGlnGlnProProGlyProArgGlyProArgGlyAspArgGly 170
 Db 151 LysAsnPheThrIleLeuGlnGlnProProGlyProArgGlyProArgGlyAspArgGly 170

Db 1498 CGCAATGTACCATCTCTACGAGGTGCCCGCCCGCCCTCCAGGACCAAGAGATTCAAAGA 1557
 QY 171 SerGlnGlyProProGlyProThrGlyAsnLysGlnLysGlyGlu----- 186
 Db 1558 GATATGGGCGTGAAGGGCTCTTGGCGCAGAGGCCGCAAGAGAGACCCCGCATCTTG 1617
 QY 187 -----LysGlyGluProGlyProGlyProGlyProGlyProGlyProGlyProGly 198
 Db 1618 GGGCCCTGGGACCCAGGGTCTCAGGGGCAACCTGGAGAGCGCGGCTGTGGAGAA 1677
 QY 199 ArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysSerLysGly 218
 Db 1678 AGGGCCCTGTGGCCCTCGAGGGTCCAGGGCTCAAAGGCTCAAAGGCTGAGCTTGA 1737
 QY 219 SerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSer 238
 Db 1738 ACTGGAGGGCGAGA-----GGACAGCCAGGCCCAAGAGGGGACATA 1779
 QY 239 GlyAspProGlyProProGlyProGlyLysGluGlyLeuProGlyProGlnGlyPro 258
 Db 1780 GGGCCCGCAGGGCCAGAGGGCCCGCGGCTCTCCAGGCGCTCAGGGCTCAGGGAAAA 1839
 QY 259 ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGly 278
 Db 1840 CCGGAATTGCAAGGAAGAGGGTCCAGGCCAGCGGGGGCCATGGGCTTAAGGT 1899
 QY 279 LeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGly 296
 Db 1900 GAACCAAGGATCCAGGGTCCCCCTGTCTCCCGGGGCTCCAGGTCCACCAAGCA 1953
 RESULT 15
 AAX27856
 ID AAX27856 standard; DNA; 3810 BP.
 XX
 AC AAX27856;
 XX
 DT 02-JUN-1999 (first entry)
 XX
 DE Human CSR1 protein coding sequence.
 XX
 KW Cellular stress response protein; CSR1; CSR2; CSR3; human; macrophage;
 KW scavenger receptor protein; intracellular stress; arteriosclerosis;
 KW diabetic circulatory obstruction; microbial infection; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9909159-A1.
 XX
 PD 25-FEB-1999.
 XX
 PF 12-AUG-1998; 98WO-JP03602.
 XX
 PR 30-JUL-1998; 98JP-0230121.
 PR 13-AUG-1997; 97JP-0233396.
 XX
 PA (NISR) JAPAN TOBACCO INC.
 XX
 PI Nakamura Y, Tokino T;
 XX
 DR WPI; 1999-181032/15.
 DR P-PSDB; AAY00992.
 XX
 PT Scavenger receptor proteins - for treatment and diagnosis of
 PT disorders involving cell stress
 XX
 PS Claim 4; Page 119-127; 175pp; Japanese.
 XX This sequence encodes the human cellular stress response 1 (CSR1) protein
 CC of the invention. The CSR proteins are macrophage scavenger receptor
 CC proteins. The CSR proteins can be used in the treatment, gene therapy
 CC and diagnosis of diseases in which intracellular stress is important,
 CC such as arteriosclerosis, diabetic circulatory obstruction, and microbial
 CC infection. Expression of the proteins is induced in vivo in response to
 CC infection.

Search completed: March 21, 2003, 08:38:02
Job time : 262.038 secs

US-09-763-712A-2_COPY_91_547 (1-457) x AAX27856 (1-3810)

Figure 1

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:26:21 ; Search time 48.552 seconds

(without alignments)

2886.622 Million cell updates/sec

Title: US-09-763-712A-2_COPY_91_547

Perfect score: 2520

Sequence: 1 MNSQNSPTGOMENTTISQ.....EDVANNICEKRETVLSSAL 457

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09763712/runat_14032003_100951_18187/app_query.fasta_1.1877
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712@cgn.1.108 @runat_14032003_100951_18187 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-FGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA:*
1: /cgn2.6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2.6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|------------------|
| 1 | 459.5 | 18.2 | 810 | 1 | US-08-642-255-60 |
| 2 | 452.5 | 18.0 | 3181 | 1 | US-08-655-086-1 |
| 3 | 448.5 | 17.8 | 756 | 1 | US-08-642-255-50 |
| 4 | 436.5 | 17.3 | 1608 | 4 | US-09-029-348-19 |
| 5 | 428 | 17.0 | 4031 | 1 | US-08-159-784-1 |
| 6 | 424.5 | 16.8 | 1572 | 4 | US-09-297-269-39 |
| 7 | 419 | 16.6 | 2543 | 3 | US-08-555-669-11 |
| 8 | 419 | 16.6 | 2543 | 3 | US-09-073-663-11 |
| 9 | 412.5 | 16.4 | 432 | 1 | US-08-642-255-48 |
| 10 | 409 | 16.2 | 3394 | 1 | US-08-159-784-4 |
| 11 | 402.5 | 16.0 | 1416 | 1 | US-07-621-091G-1 |
| 12 | 402.5 | 16.0 | 1416 | 2 | US-08-399-889-1 |

| | | | | | | |
|----|-------|------|-------|---|--------------------|--------------------|
| 13 | 402.5 | 16.0 | 1416 | 3 | US-09-167-364-1 | Sequence 1, Appli |
| 14 | 402.5 | 16.0 | 1416 | 4 | US-09-439-897-1 | Sequence 1, Appli |
| 15 | 400.5 | 15.9 | 4359 | 4 | US-09-484-970B-4 | Sequence 4, Appli |
| 16 | 399 | 15.8 | 1881 | 4 | US-09-029-348-20 | Sequence 20, Appli |
| 17 | 394.5 | 15.7 | 1560 | 2 | US-08-794-795-5 | Sequence 5, Appli |
| 18 | 394.5 | 15.7 | 1560 | 4 | US-09-249-200-5 | Sequence 5, Appli |
| 19 | 394 | 15.6 | 5102 | 1 | US-08-494-168-1 | Sequence 40, Appli |
| 20 | 393.5 | 15.6 | 503 | 4 | US-09-297-269-40 | Sequence 1, Appli |
| 21 | 391 | 15.5 | 1560 | 4 | US-09-453-702B-264 | Sequence 264, App |
| 22 | 391 | 15.5 | 61663 | 4 | US-09-453-702B-62 | Sequence 62, Appli |
| 23 | 389.5 | 15.5 | 1703 | 2 | US-08-794-795-1 | Sequence 1, Appli |
| 24 | 389.5 | 15.5 | 1703 | 4 | US-09-249-200-1 | Sequence 1, Appli |
| 25 | 380.5 | 15.1 | 1868 | 3 | US-08-392-367B-1 | Sequence 1, Appli |
| 26 | 380.5 | 15.1 | 1868 | 3 | US-08-893-467A-1 | Sequence 1, Appli |
| 27 | 356 | 14.1 | 1839 | 1 | US-08-383-744-1 | Sequence 1, Appli |
| 28 | 356 | 14.1 | 1839 | 2 | US-08-999-336-1 | Sequence 1, Appli |
| 29 | 356 | 14.1 | 1839 | 5 | PCT-US96-01427-1 | Sequence 1, Appli |
| 30 | 355.5 | 14.1 | 2383 | 4 | US-09-523-487-9 | Sequence 9, Appli |
| 31 | 355.5 | 14.1 | 2409 | 3 | US-09-320-095-9 | Sequence 9, Appli |
| 32 | 350 | 13.9 | 1458 | 4 | US-09-111-470-3 | Sequence 3, Appli |
| 33 | 343 | 13.6 | 506 | 1 | US-08-253-155A-61 | Sequence 61, Appli |
| 34 | 331 | 13.1 | 2824 | 2 | US-09-010-928B-3 | Sequence 3, Appli |
| 35 | 330.5 | 13.1 | 392 | 4 | US-09-404-879A-372 | Sequence 372, App |
| 36 | 326.5 | 13.0 | 48908 | 4 | US-09-453-702B-137 | Sequence 137, App |
| 37 | 325.5 | 12.9 | 9827 | 4 | US-09-453-702B-66 | Sequence 66, Appli |
| 38 | 323.5 | 12.8 | 3404 | 4 | US-09-453-702B-94 | Sequence 94, Appli |
| 39 | 322 | 12.8 | 1370 | 4 | US-09-111-470-9 | Sequence 9, Appli |
| 40 | 320.5 | 12.7 | 45175 | 4 | US-09-453-702B-116 | Sequence 116, App |
| 41 | 315.5 | 12.5 | 2851 | 4 | US-09-535-521-1 | Sequence 1, Appli |
| 42 | 315.5 | 12.5 | 2851 | 4 | US-09-535-521-3 | Sequence 3, Appli |
| 43 | 310.5 | 12.3 | 876 | 4 | US-09-535-521-4 | Sequence 4, Appli |
| 44 | 310.5 | 12.3 | 876 | 4 | US-09-535-521-6 | Sequence 6, Appli |
| 45 | 306.5 | 12.2 | 38584 | 4 | US-09-453-702B-50 | Sequence 50, Appli |

ALIGNMENTS

RESULT 1

US-08-642-255-60
; Sequence 60, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:

```

:
: COUNTRY: US
: ZIP: 11553
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0. Version #1.25
:

```

```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US/08/655,086
3 FILING DATE: 03-JUN-1996
4 CLASSIFICATION: 435
5 ATTORNEY/AGENT INFORMATION:
6 NAME: STEEN, JEFFREY S.
7 REGISTRATION NUMBER: 32,063
8 REFERENCE/DOCKET NUMBER: 203-1632

```

TELECOMMUNICATION INFORMATION;
TELEPHONE: 516-228-8484

```

/ BLASTX: J16228.6310
/
/ INFORMATION FOR SEQ ID NO: 1:
/
/ SEQUENCE CHARACTERISTICS:
/   LENGTH: 3181 base pairs
/   TYPE: nucleic acid
/   STRANDEDNESS: single
/   TOPOLOGY: unknown
/
/ MOLECULE TYPE: cDNA
/

```

Alignment Scores: 3.58e-17 Length: 3181
Pred. No.:

| | | | |
|------------------------|--------|---------------|----|
| Score: | 452.50 | Matches: | 33 |
| Percent Similarity: | 52.11% | Conservative: | 12 |
| Best Local Similarity: | 46.48% | Mismatches: | 53 |
| Query Match: | 17.96% | Indels: | 49 |

US-09-763-712A-2_COPY_91_547 (1-457) x US-08-655-086-1 (1-3181)

QY 156 leuGlnclYPrIoPrgYPrIoAGlGlyPrIoAGlGlyAspArgGlySeIcGlnclYPrIoFIO 173
:::|||||||||| 111 ||| |||||:::||||| |||
Db 893 GTTCAGGACCCCTGGCCCTGCTGGAGAGGAAGCAAGCGAGGAGCTCGAGGTGAACCC 952

QY 1/6 GYGFPRFNRGTYASNLGSGYGNLYGSLGYNLYSGYGLYGRUPTG..... 190
||||| ||||| ||| |||||: ||| |||
Db 953 GGACCCACTGGCGCTGCCGGGACCCCTGGCGAGCGTGGTGGACCTGGTAGCGGTGTTTC 1012

| | | | |
|----|------|--|------|
| | I91 | - - - - - GlyProPoglyProAlagIguargelyProIIlegly | 203 |
| Qy | | | |
| Dd | 1013 | CCTGGCCAGATGgTGTGtTGtGCCcCAGAggGCccCGctGGtgAACgtGGTCtcTTCCtgcC | 1072 |

QY 204 ProAlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
||||| 111111111 111111111 111111111 111111111
Db 1073 CCCGCGGCCCCAAAGGATCTCCGTGGTAAGCTGGTCTCCCGGTGAAGCTGGTCTGCCCT 1132

QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
||| ||||| ||||||||| ||||||| ||| :::
Db 1133 GGTGCCAAGGGTCTGACTGGAAAGCCCTGGCAGCCCTGGTCCTGATGGCAAACT----- 1186

| | | | |
|----|------|---|------|
| Qy | 241 | ProGlyProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProPro--- | 259 |
| Db | 1187 | ---GGCCCCCCTGTCCCGCGGTCAAGATGTCGCCCCCGACCCCAGGCCACCCTGGT | 1243 |

| | | | |
|----|------|---|------|
| Qy | 260 | -----GlyPheGlnGlyLeuGlnGlyThrValGlyGluPro | 271 |
| | | :: | |
| Db | 1244 | GCCCGTGGTCAGGCTGGTGTGATGGGATTTCCCTGGACCTAAAGGTGCTGCTGGAGAGCCC | 1303 |

| | | | | |
|----|------|--|-------|------|
| Qy | 272 | GlyValProGlyProArgGlyLeuProGlyLeuProGlyVal | ----- | 285 |
| | | :: | | |
| Db | 1304 | GGCAAGSCTGGAGACGAGGCTTCCCGACCCCTGGCGCTGTGGTCTGCTGGCAAA | | 1363 |

| | | | |
|----|------|--|------|
| Qy | 286 | ProGlyMetProGlyProLysGlyProGlyProGlyProGlyProSerGly----- | 302 |
| | | :: :: | |
| Db | 1364 | GATGGAGAGGCTGGAGCTCAGGAGACCCCTGGCCCTGCTGTCCTGGCGAGAGAGGT | 1423 |

```

QY 303 -----AlaValValProLeuAlaLeuGlnAsnGluProThrProAla----- 316
Db 1424 GAACAAGGCCCTGCTGGCTCCCC-----GGATTCCAGGGTCTCCCTGGTCTCT 1480

QY 317 -----ProGluAspAsnGlyCysPro 323
Db 1481 CCAGGTGAAGCAGGCAACCTGGTGAACAGGGTGTCTCT 1519

RESULT 3
US-08-642-255-50
; Sequence 50, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A5556-3/BI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 756 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic"
US-08-642-255-50

Alignment Scores:
Pred. No.: 1.35e-17 Length: 756
Score: 448.50 Matches: 99
Percent Similarity: 52.70% Conservative: 18
Best Local Similarity: 44.50% Mismatches: 85
Query Match: 17.80% Indels: 21
DB: 1 Gaps: 5

US-09-763-712a-2_COPY_91_547 (1-457) x US-08-642-255-50 (1-756)
QY 107 IleArgLeuAspSerValSerLeuArgMetGlnAspLeuMet-----ArgSerArg 124
Db 56 GTGGCGCTGACCGGCTGTCTCCACCGGCTGCTCCGGGACCTGCAGGCCCGCCAGGTGCGC 115

QY 125 LeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLysLeuValAspSer 144
Db 116 CTGGACCGGCTGTCTCCACCGG-----GTGCTCCGGGACCTGCAGGCCCGCCAGGTG 166

QY 145 LysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProProGlyProArgGly 164
Db 167 CGCCTGGACCGGCTGTCTCCACCGG-----GTGCTCCG--GGACCTGCAGGCCCGCCAGGT 219

```

```

QY 165 ProArgGlyAspArgSerGlnGlyProProGlyProThrGlyAsnLysGlyGlnLys 184
Db 220 GGCCTGGACCGGCTGTCTCCACCGGCTGTCTCCGGACCTGCAGGCCCGCCAGGTGCGCCT 279

QY 185 GlyGluLysGlyGluProProGlyProProGlyProAlaGlyGluArgGlyProIleGlyPro 204
Db 280 GGACCGGCTGTCTCCACCGGCTGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCG 339

QY 205 AlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGly 224
Db 340 GCTGTGTCCACCGGCTGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCGGCTGTGT 399

QY 225 SerArgGlySerPro-----GlyLysProGlnGlyProGlnGlyProSerGlyAspPro 241
Db 400 CCACCGGCTGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCGGCTGTGTCCACCG 459

QY 242 GlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProProGlyPhe 261
Db 460 GGTGTCTCCGGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCGGCTGTGTCCACCGGCTGTGT 519

QY 262 GlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuProGly 281
Db 520 CCGGGACCTGCAGGCCCGCCAGGTGCGCCTGGACCGGCTGTGTCCACCGGCTGTGTCCGGGA 579

QY 282 LeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyProSer 301
Db 580 CTTGCAGGCCCGCCAGGTGCGCCTGGACCGGCTGTGTCCACCGGCTGTGTCCGGGACCTGTGA 639

QY 302 GlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGly 321
Db 640 GGC-----CCGCCAGGTGCGCCTGGACCGGCTGTGT 669

QY 322 CysPro 323
Db 670 CCACCG 675

RESULT 4
US-09-029-348-19
; Sequence 19, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: 087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1608
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-19

Alignment Scores:
Pred. No.: 1.38e-16 Length: 1608
Score: 436.50 Matches: 102
Percent Similarity: 47.06% Conservative: 18
Best Local Similarity: 40.00% Mismatches: 67
Query Match: 17.32% Indels: 70
DB: 4 Gaps: 9

US-09-763-712a-2_COPY_91_547 (1-457) x US-09-029-348-19 (1-1608)
QY 145 LysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProProGlyProArgGly 164
Db 59 CAACATGCCCACTTTACAGAGGAA--ACTGTAAGAAAGAGGCCCGCCAGGTATAGAGA 117

QY 165 ProArgGlyAspArgSerGlnGlyProProGlyProThrGlyAsnLysGlyGlnLys 184

```

Db 118 CCACGTGGAGAGGGTCCACAGGCCCCCCCCAGGACAGATGGTGAAGATGGTCCACA 177
QY 185 GlyGluLysGlyGluProGlyProGlyPro----- 195
Db 178 GGCCCTCCTGGTCCACCTGGTCTCTCTGCCCCCTGGTCTGGTGGGACATTTGGTGT 237
QY 196 -----AlaGlyGluArgGlyProLeuGlyProAla 205
Db 238 CAGTATGATGAAAGAGTGGACTTGGCCCTGGACCAATGGCTTAATGGACCTAGA 297
QY 206 GlyProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySer 225
Db 298 GGCCCACTGGTGCAGCTGGAGCCCCAGGCCCTCAAGGTTTCCAAAGGACCTGCTGTGAG 357
QY 226 ArgGlySerProGlyLys-----ProGlyProGlnGlyPro 237
Db 358 CTGGTGAACCTGGTCAAACTGGTCTGCAGGTGCACCTGGTCTCATGCCCCCTGGGT 417
QY 238 -----SerGlyAspProGlyProGlyProGlyProGlyLysGlu 250
Db 418 CTGCTGGCAACATGAAACCGTGTGAACCTGGTCTCTGGTCTCTGGTCTCTGGT 477
QY 251 GlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGlu 270
Db 478 GGTGCTGTGGCCCAAGAGGTCTAGTGGCCCAAGGACATTCGTGGCGATAAGGAGAG 537
QY 271 ProGlyValProGlyProArgGlyLeuPro-----Gly 281
Db 538 CCGGGTAAAGGGGCCAGAGGTCTCTCTGGTTCGAAGGACACAAATGGATTGCAAGGT 597
QY 282 LeuProGlyVal-----ProGlyMetProGlyProLys 292
Db 598 CTGCCTGGTATCGTGTGTCCACATGTGTATCAAGGTGCTCTGGTCTCGTGGTCTGCT 657
QY 293 GlyProGlyProGlyProGlyProSerGly----- 302
Db 658 GGTCTTAGGGGGCCCTGCTGTCTCTGGCCCTGGTGGAAAGATGGTGCACATGGACAT 717
QY 303 ---AlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGly 321
Db 718 CTGGTACGGTGGACCTGCTGG-CATTGAGGCCCTCAGGTGCACCAAGGCC----- 770
QY 322 Cys---ProPro-----HisTrpLysAsnPheThrAspLysCys 333
Db 771 TGTGGCCCCCTGGTCCCTGGCCCTCCTGGACCTCCAGGTGT 815

RESULT 5

US-08-159-784-1
; Sequence 1, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4031
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-159-784-1
Alignment Scores:
Pred. No.: 1.07e-15 Length: 4031
Score: 428.00 Matches: 95
Percent Similarity: 46.41% Conservative: 15
Best Local Similarity: 40.08% Mismatches: 49
Query Match: 16.98% Indels: 78
DB: 1 Gaps: 8
US-09-763-712a-2_COPY_91_547 (1-457) x US-08-159-784-1 (1-4031)
QY 141 LeuValAspSerLysHisGlyGlnLeuLeuLysAsnPheThrIleLeuGlnGlyProPro 160
Db 976 GTGGTCCAGGCCCCCAACTCAACAACCTGTCCCTGGA-----GCACAAGGCCCCCG 1026
QY 161 GlyProArgGlyPro----- 165
Db 1027 GGACCTCAGGGGCCACCGAGGAGGATGCCACTCCAGGAAGGATGGTGAACGGGTGAC 1086
QY 166 -----ArgGlyAspArgGlySerGlnGlyProProGlyProThrGly 179
Db 1087 CCGTGTGAAGATGGGAGACCGGGTGTGACACTGACCTCAAGGCTTTCAGGGACCCCGAG 1146
QY 180 AsnLysGlyGlnLysGlyGluLysGlyGluPro-----GlyProProGly 194
Db 1147 GATGTGGCCCTCAAGGGCGAGGAGATCCTGTATTGGGCCCGCCGAGACCTCCAGGG 1206
QY 195 ProAlaGlyGluArgGlyProLeu----- 202
Db 1207 CCTCCAGGGCCACCGAGGACCTCTTTCACACAAGACAAGTGCACCTTCATTGACATGGAG 1266
QY 203 -----GlyPro 204
Db 1267 GGATCCGGTTTCAGCGGAGACATAGAGAGCCTTAGAGGCCACGAGGCTTCCCTGGCCCC 1326
QY 205 AlaGlyPro-----ProGlyGluArgGlyGlyLysGlySerLysGly 218
Db 1327 CCGGGCCCCCTGGTCTCCAGGACTTCTGTGTGGCCAGGACGCTTGGGATCAATGTT 1386
QY 219 SerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSer 238
Db 1387 TCCATGCACCA--GGACCTGCAGGCCCTCTCTGGTGTACCTGGGAAGGAGGCCCC 1443
QY 239 GlyAspProGlyProGlyProGlyLysGlyGluLeuProGlyProGlnGlyPro 258
Db 1444 GGTTCCTCAGGTCCCGGGGACCTCCAGGTCT-----CCAGGCAAGAGGCCCA 1494
QY 259 ProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArg--- 277
Db 1495 CCAGGAGTGGCCCGCAGAAAGGCGAGTGTGTGTGGTGTGGCATCCAGGACCCAGGGG 1554
QY 278 -----GlyLeuProGlyLeuProGlyValProGly 289
Db 1555 AGCAAGGAGACCTTGGGCCCATCGTATGCTGTGGCAAGTCTGGCTGTGGTGTGATCCCT 1614
QY 290 GlyProLysGlyProGlyProGlyProGlyProGlyProGlyProGlyProGlyPro 306
Db 1614

Db 1615 GGCCAGCTTGGACCCCGGAGGAGCTCCAGGCGCTCCAGGCGCCAGGAGCA 1665

RESULT 6

US-09-297-269-39
 ; Sequence 39, Application US/09297269
 ; Patent No. 6451557
 ; GENERAL INFORMATION:
 ; APPLICANT: VAUGHAN, Paul R.
 ; APPLICANT: RAMSHAW, John A.M.
 ; APPLICANT: WERKMEISTER, Jerome A.
 ; TITLE OF INVENTION: A METHOD FOR PRODUCING, IN YEAST, A HYDROXYLATED TRIPLE
 ; HELICAL PROTEIN, AND YEAST HOST CELLS USEFUL IN SAID
 ; TITLE OF INVENTION: METHOD (As Amended)
 ; FILE REFERENCE: Q54094
 ; CURRENT APPLICATION NUMBER: US/09/297,269
 ; CURRENT FILING DATE: 1999-04-28
 ; EARLIER APPLICATION NUMBER: P03310
 ; EARLIER FILING DATE: 1996-10-29
 ; EARLIER APPLICATION NUMBER: P04306
 ; EARLIER FILING DATE: 1996-12-19
 ; EARLIER APPLICATION NUMBER: PCT/AU97/00721
 ; EARLIER FILING DATE: 1997-10-29
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 39
 ; LENGTH: 1572
 ; TYPE: DNA
 ; ORGANISM: Collagen Type III Alpha I Chain
 US-09-297-269-39

Alignment Scores:

Pred. No.: 6,33e-16 Length: 1572
 Score: 424.50 Matches: 119
 Percent Similarity: 42.48% Conservative: 25
 Best Local Similarity: 35.10% Mismatches: 108
 Query Match: 16.85% Indels: 88
 DB: 4 Gaps: 15

US-09-763-712A-2_COPY_91_547 (1-457) x US-09-297-269-39 (1-1572)

QY 134 ValIleMetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPro 153
 Db 194 GTAACTGGAGAGATGGAACCCCTGGAT-----CAGATGGTC-----TTC 235
 QY 154 ThrIleLeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySer----- 171
 Db 236 CAGGTGCA-GATGGATCTCTGCTGTCGCAAGGGTGATCGTGGTGAAATGGCTCTCTCGGT 294
 QY 172 -----GlnGlyProGlyProGlyProThrGlyAsnLysGlyGlnLys 184
 Db 295 GCGCCTGGCGCTCTGCTGTCATCCAGCCACCTGGTCTGCTGCTGCGTGGAAAGAGT 354
 QY 185 GlyGluLysGlyGluProGlyPro-----ProGlyProAlaGlyGlu 198
 Db 355 GTGACAGAGGAGAGAGAGTGGCCCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 414
 QY 199 ArgGlyProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGly 218
 Db 415 CGAGGTGCTCTGCTCTCAAGGCCCCAGTGGTGACAAAGGTGAACAGGTGAACGTGGA 474
 QY 219 SerGlnGlyProLysGlySerArg-----GlySerProGlyLysPro 232
 Db 475 GCTGCTGGCATCAAGAGCATCAGAGATTCCTGGTAATCCAGGTGCCCGGCTTCTCCA 534
 QY 233 GlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlyLys 252
 Db 535 GGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 594
 QY 253 ProGlyProGlnGlyProGlyProGlyProGlyProGlyProGlyProGlyProGly 272
 Db 595 GTTGGACCCAGTGACCTCTCTGCAAGAGATGGAACCAAGTGGACATCCAGGTCCCATTTGA 654

QY 273 ValProGlyPro-----ArgGlyLeuProGlyLeuProGlyValPro 286
 Db 655 CCACCAAGGCGCTCGAGGTAAACAGAGGTGAAAGAGGATCTCGAGGCTCCCGCAGGCCCA 714
 QY 287 GlyMetProGlyProGlyProGlyProGlyProGlyProGlyProSer-----Gly 302
 Db 715 GGGCAACAGGCGCTCTGAGCTCTGAGCTCTGAGTGGTGGTGGTGGTGGTGGTGGTGGT 774
 QY 303 AlaValProLeuAlaLeuGlnAsnGlu-----ProThrProAla 316
 Db 775 GCGGCTGCATTCCTGGGATGGAGGTGAAAGAGTGGCGGTTTGGCCCGTATTATGGA 834
 QY 317 ProGluAspAsnGlyCysProHisTrpLysAsnPhetrAsp----- 331
 Db 835 CTTGAA-----CCAATGGATTTCAAAATCAACACCGATGAGATTACACTTCA 882
 QY 332 ---LysCysTyrTrpPheSerValGluLysGluIlePheGluAspAlaLys----- 347
 Db 883 CTCAGTCTGTTAATAGCAAAATAGAAAGCCTCATTAGTCTGCTGCTGCTGCTGCTGCT 942
 QY 348 -----LeuPheCysGluAspLys----- 353
 Db 943 CCGCTAGAAACTGCAGAGACCTGAAATCTGCCATCCTGAACTCAAGACTGGAGATAC 1002
 QY 354 -----SerSerHisLeuValPheIleAsnThrArg 363
 Db 1003 TGGGTGCGACCCCTAACCAAGGATGCAAAATGGATGCTATCAAGTATTCTCTAATATGGA 1062
 QY 364 GluGluGlnGlnTrpIleLysLysGlnMetValGly---ArgGluSerHisTrpIleGly 382
 Db 1063 ACTGGGGAACATGCATAAGTGCCTTGAATCTTCCACGAAACACTGG----- 1116
 QY 383 LeuThrAspSerGluArgGluAsnGluTrpLysTrp-----LeuAspGly 397
 Db 1117 TGGACAGATCTAGTCTGCTGAGAGAAACACCGTTTGGTTTGGAGAGTCCATCGATG 1173
 RESULT 7
 US-08-555-669-11
 ; Sequence 11, Application US/08555669
 ; Patent No. 5773248
 ; GENERAL INFORMATION:
 ; APPLICANT: Brewton, Richard G.
 ; APPLICANT: Mayne, Richard
 ; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
 ; NUMBER OF SEQUENCES: 32
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/555,669
 ; FILING DATE: 13-NOV-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Halluin, Albert P.
 ; REGISTRATION NUMBER: 25,227
 ; REFERENCE/DOCKET NUMBER: 8389-030
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-854-3660
 ; TELEFAX: 415-854-3694
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2543 base pairs
 ; TYPE: nucleic acid

```
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..2098
US-08-555-669-11

Alignment Scores:
Pred. No.: 2,11e-15 Length: 2543
Score: 419.00 Matches: 121
Percent Similarity: 35.59% Conservative: 26
Best Local Similarity: 29.30% Mismatches: 91
Query Match: 16.63% Indels: 176
DB: 1 Gaps: 9

US-09-763-712a-2_COPY_91_547 (1-457) x US-08-555-669-11 (1-2543)
QY 142 ValAspSerLysHisGlnLeuIleLysAsnPheThrIleLeuGlnGlyProProGly 161
DB 986 CTCGATGGCCAGGAAGGAGAGGCTGGTCCGAAC-----GGTGTCTCGGGA 1030
QY 162 ProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGlyAsnLys 181
DB 1031 GAGAAGGGCCCCAACGGGCTCCCGGCTCCCTGGAGAGGGGGTCCAAAGCGCAGAAG 1090
QY 182 GlyGlnLysGlyGlnLysGlyGluProGlyProProGlyProAlaGlyGluArgGlyPro 201
DB 1091 GGAGAACGGGGCAGAGCTGGGGAGCTGGGTGAGGCGCGGCCCTCTGGAGAGCCAGCGCTC 1150
QY 202 IleGlyProAlaGlyProProGlyGluArgGlyGlySerLysGlySer----- 219
DB 1151 CTTGGAGATGCTGGCATGCTGGGGAGCGCGGTGAGGCTGGCCACCGGGCTCAGCGGGG 1210
QY 220 -----GlnGlyProLys 223
DB 1211 GCCTCGGCCCAAGCCCTCCCGAGGCCCTGGTGTCTGGAGGCTTCAGGCGCCAGAAG 1270
QY 224 GlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyPro 243
DB 1271 GGCAGCATGGGAGACCCCGGCTCTCCAGGCCCCCAGGGCTCCGAGGTGACGTGGCGCAC 1330
QY 244 ProGlyProGlyLysGlnGlyLeuProGlyProGlnGlyProProGlyPheGlnGly 263
DB 1331 CGGGGTCCGGGAGTGCCTGGAAGGCCCTAAAGGAGACACAGGGTATTGACGGTTCACGGT 1390
QY 264 LeuGlnGlyThrValGlnGly----- 270
DB 1391 CTTCTGGGGATAAAGAGAGACTGGGTCCAGCGGCTGGTCGAGCCCAAGAGAGATCT 1450
QY 270 ----- 270
DB 1451 GGCAGTCGAGGGAGCTGGGCCCCCAAGGACCCAGGGTCCCAACGGCACCCAGCGGTGT 1510
QY 271 -----ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly 287
DB 1511 CAGGGTGTCCCGGGCCCCCGGCTCTCTGGGCTCGAGGGGCTCCCGGGGTCTCTGCTGGC 1570
QY 288 Met----- 288
DB 1571 ATCAGCGGGAAGCGGGAGTTCGGGGAGGAGGAGCGCAGCAGCGCATCAGGGAGCTG 1630
QY 288 ----- 288
DB 1631 TGTGGGGGATGATCAGCAACAAATTCACAGTTAGCGCGGCACCACTAAGGAAGCCTTTG 1690
QY 289 -----ProGlyProLysGlyProGlyProProGlyProSer 301
DB 1691 GCACCGGGTCCATTGGTCGGCGCGGTCCAGGTGGGCCCCCTGGGCCCCCAGGACCCCCA 1750
QY 302 GlyAlaVal-ValProLeuAlaLeuGlnAsnGluPro----- 313
DB 1751 GGCCTCATTTGGTCACCTTGGCGCTCGAGGACCCCCCGGATACCGCGGTCTCCACTGGGGAG 1810
```

```
QY 314 -----ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAs 331
DB 1811 CTGGGAGACCCCGGCCAGAG----- 1832
QY 331 pLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysG1 351
DB 1833 -----GAAACCAAGGTGCACAGAGAGACAAG----- 1859
QY 351 uAspLysSerSerHisLeuValPheIleAsnThrArgGluGlnGlnTrpIleLysLy 371
DB 1860 -----GCGCGGCAGAGCAG----- 1874
QY 371 sGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnG1 391
DB 1875 -----GGCTGGACGGCTCGAAGAGACACAGGCCCCCAAGGA 1912
QY 391 uTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAs 411
DB 1913 CCCCCAGGC-GTCCCGGCACCAAGAGGACGCGGCGGCTGCTCCCGCGAGCCTGG 1971
QY 411 pAsnTrpGlyHisGlyHisGlyProGly----- 420
DB 1972 GCCTCCCGGAGATCTCGGGCTTCAGGTGCCATTGGGCCCCAGGACACCGGGGATCTG 2031
QY 421 -----GluAspCysAlaGlyLeuIleTyrAlaGly 430
DB 2032 CGACACCTCAGCCTGCCAAGGAGCGCGTGTAGGAGGG 2068

RESULT 8
US-09-073-663-11
; Sequence 11, Application US/09073663
; Patent No. 6127523
; GENERAL INFORMATION:
; APPLICANT: Brewton, Richard G.
; TITLE OF INVENTION: TYPE IX COLLAGEN AND FRAGMENTS THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McGregor & Adler, LLP
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS 8.1
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,663
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benjamin Aaron Adler, Ph.D., J.D.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5913D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 777-2321
; TELEFAX: (713) 777-6908
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2543 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..2098
US-09-073-663-11
```


| | | | |
|----|------|---|------|
| Qy | 142 | ValAspSerLysHisGlyGlnLeuLeuLysAsnPheThrIleLeuGlnGlyProProGly | 161 |
| Db | 986 | CTCGATGGCCAGAAAGGAGAGAGCTGGTCCGAAC-----GGTGTCTCCGGGA | 1030 |
| Qy | 162 | ProArgGlyProCArgGlyAspArgGlySerGlnGlyProProGlyProThrGlyAsnLys | 181 |
| Db | 1031 | GAGAAAGGCCCAAGCGGCTGCCGGGCTCTCGTGACGAGCGGGGTCCAAAGGCGAGAG | 1090 |
| Qy | 182 | GlyGlnLysGlyGluLysGlyProGlyProGlyProAlaGlyGluArgGlyPro | 201 |
| Db | 1091 | GGAGAACGGGCGACAGCTGGGGAGCTGGTGAGGCGCGCCCTCTGGAGACGACGGGCTC | 1150 |
| Qy | 202 | IleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySer----- | 219 |
| Db | 1151 | CCTGGAGATGCTGGCATGCTCTGGGAGGCGGTGAGGCTGGCCACCGGGCTCAGCGGG | 1210 |
| Qy | 220 | -----GlnGlyProLys | 223 |
| Db | 1211 | GCCTCGGCCACAGGCGCTCCCGAGCGCCGTGTGTCGAGGCTTCCAGGGCCAGAG | 1270 |
| Qy | 224 | GlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAspProGlyPro | 243 |
| Db | 1271 | GGCAGCATGGAGACCCCGGCTCTCCAGGCCCCAGGCGCTCCGAGGTGACGTGGGGCAC | 1330 |
| Qy | 244 | ProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProProGlyPheGlnGly | 263 |
| Db | 1331 | CGGGGTCCGGGAGGTGCCGAAGCGCCTAAGGGAGACACAGGATATTGCAGGTTCCGACG | 1390 |
| Qy | 264 | LeuGlnGlyThrValGlyGlu----- | 270 |
| Db | 1391 | CTTCTCGGGGATAAAGGAGAACTGGGTCCCAACGCGCCTGTCGACCCCAAGGAGAGTCT | 1450 |
| Qy | 270 | ----- | 270 |
| Db | 1451 | GGCAGTCGAGGGGAGCTGGGCCCCAAAGGCACCCAGGCTCCACAGCGCACCCAGCGGT | 1510 |
| Qy | 271 | -----ProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGly | 287 |
| Db | 1511 | CAGGGTGTCCCGGGCCCCCGGTCTCTGGGCTGTCAGGCGCTGCCGGGTGTCTCTGGC | 1570 |
| Qy | 288 | Met----- | 288 |
| Db | 1571 | ATCACGGGGAACCGGGAGTTCCGGGGAGGAGGCCACGCGACGCGCATCAGGAGCTG | 1630 |
| Qy | 288 | ----- | 288 |
| Db | 1631 | TGTGGGGGATGATCAGCGCAACAATTGCACAGTTAGCCGGCGCACCTAAGGAACCTTTG | 1690 |
| Qy | 289 | -----ProGlyProLysGlyProProGlyProProGlyProSer | 301 |
| Db | 1691 | GCACCGGGTCCATGTTGGTCCCGCTCAGCTGSCCCCTGGGCCCCCGAGGCCCA | 1750 |
| Qy | 302 | GlyAlaVal-ValProLeuAlaLeuLysGluPro----- | 313 |
| Db | 1751 | GGCTCATGGTACCCCTGGCGCTCGAGACCCCGCGATACCAGCGTCCCACTGGGGAG | 1810 |
| Qy | 314 | -----ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAs | 331 |
| Db | 1811 | CTGGGAGACCCGGGCCAGAG----- | 1832 |
| Qy | 331 | pLysCysTyrtyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysG1 | 351 |
| Db | 1833 | -----GAAACCAAGGTGCACAGGAGACAAG----- | 1859 |

US-09-763-712A-2_COPY_91_547 (1-457) x US-08-642-255-48 (1-432)

```
QY 158 GlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyPro 177
    ||||| |||||:|||| ||| :||| |||||
Db 10 GCGCCAGCAGGTCGGAAGGCGCGATGCGCCAGCAGGCGCCGAAAGGTGCGCTGGACCG 69
QY 178 ThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGly 197
    ||| ||| ||| ||| ||||| ||||| ||||| |||||
Db 70 GCTGGTCCACCGGGTGCTCCGGGACCTGCAGGCGCGCCAGGTGCGCTGGACCGGTGT 129
QY 198 GluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLys 217
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 CCACCGGGTGCTCCGGGACCTGCAGGCGCGCCAGGTGCGCTGGACCGGTGCTCCACCG 189
QY 218 GlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyPro 237
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 GGTGCTCCGGGACCT-----GCAGGCGCGCCAGGTGCGCTGGACCGGTGCTGCTCA 240
QY 238 SerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGly 257
    ||| ||| ||||| ||||| ||||| ||||| |||||
Db 241 CCGGGTGCTCCGGGACCTGCAGGCGCGCC-----GGTGGCTGGACCGGTGT 291
QY 258 ProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArg 277
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 292 CCACCG-----GGTGCTCCGGGACCTGCAGGCGCGCCAGGTGCGCTGGACCGGT 342
QY 278 GlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyPro 297
    ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 343 GGTCCACCGGGTGCTCCGGGACCTGCAGGCGCGCCAGGTGCGCTGGACCGGTGCGCG 402
QY 298 ProGlyProSerGlyAlaValPro 306
    ||| ||| |||
Db 403 AAAGAGCTCACGGTCCGGCAGGTCCG 429
```

RESULT 10

```
US-08-159-784-4
; Sequence 4, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
```

```
COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55sx
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,784
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: John F. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/170001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3394
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-159-784-4
Alignment Scores:
Pred. No.: 1.03e-14 Length: 3394
Score: 409.00 Matches: 86
Percent Similarity: 48.98% Conservative: 10
Best Local Similarity: 43.88% Mismatches: 42
Query Match: 16.23% Indels: 58
DB: 1 Gaps: 7
US-09-763-712A-2_COPY_91_547 (1-457) x US-08-159-784-4 (1-3394)
QY 156 LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySer----- 171
    ||||| ||||| ||||| ||||| ||||| |||||
Db 553 ATGCCCGCGCGCCAGGACCTCCAGGCGCGCCAGGCGCTCCAGGACCTCTCTTTAGGAC 612
QY 172 -----GlnGlyProProGlyProGlyProGlyProGlyProGlyProGlyPro 182
    ||||| ||||| ||||| ||||| ||||| |||||
Db 613 AGCAATGTGTTTGTCTGAGTCCAGCGCGCGCGCGCTCCAGGATTGCCAGGAATCAG--- 669
QY 183 GlnLysGlyGluLysGlyGluProGlyProProGlyProAlaGlyGluArgGlyPro 202
    ||||| ||||| ||||| ||||| ||||| |||||
Db 670 -----GGCCTCCAGGACCCCAAGGCGCGCCCAAGAGAGAGTG 705
QY 203 GlyProAlaGlyProProGlyGlu----- 210
    ||||| ||||| ||||| ||||| ||||| |||||
Db 706 GSCCGCGCGCGGACCAAGGCGAGTTCGGTTTGACTTTTTCAGAAGGAGGCTGAAATG 765
QY 211 ArgGlyGlyLysGlySerGlnGlyProLysGlySerArgGlySer----- 228
    ||||| ||||| ||||| ||||| ||||| |||||
Db 766 AAGGGGAGAGGAGGACCGAGGTGATGCAGGACAGAAAGGCGAGGGGAGGCGCGG 825
QY 229 -----ProGlyLysProGlyProGlnGlyProSerGly 239
    ||||| ||||| ||||| ||||| ||||| |||||
Db 826 GCGCGCGGTTCCTCGGCTCCAGCGCTGCCGGCGCGCGCGCGCGCGCGCGCGCGCG 885
QY 240 AspProGlyProProGlyProGlyLysGlu-----GlyLeuProGly 254
    ||||| ||||| ||||| ||||| ||||| |||||
Db 886 TACCCTGGGATTCAGGTCCCAAGGAGAGAGCATCGGGGCGCAGCGCGCGCGCGCGCG 945
QY 255 ProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValPro 274
    ||||| ||||| ||||| ||||| ||||| |||||
Db 946 CCTCAGGACCGCGCGCATC---GGCTACAGAGGCGCGCGCGCGCGCGCGCGCGCG 1002
QY 275 GlyProArgGlyLeuProGlyLeuProGly-----ValProGly 287
    ||||| ||||| ||||| ||||| ||||| |||||
Db 1003 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1062
QY 288 MetProGlyProLysGlyProProGlyProProGlyProProGlyProSerGlyAla 303
    ||||| ||||| ||||| ||||| ||||| |||||
Db 1063 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1110
```

RESULT 11

```
US-07-621-091G-1
; Sequence 1, Application US/07621091G
; Patent No. 5424408
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T., Morrison, Karen E., Hudson, Billy
; APPLICANT: G.
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University, Office of Cooperative Research
; STREET: 246 Church Street
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: U.S.A.
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800K storage
```

COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh OS7.0
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/621.091G
FILING DATE: 11/30/90
CLASSIFICATION: 424
PRIOR APPLICATION DATA: No. 5424408 applicable
ATTORNEY/AGENT INFORMATION:
NAME: Barth, Richard S.
REGISTRATION NUMBER: 28180
REFERENCE/DOCKET NUMBER: 900983/RB
TELEPHONE: (212) 972-1400
TELEFAX: (212) 370-1622
TELEX: 236268
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1416 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Calf
STRAIN: Unknown
INDIVIDUAL ISOLATE: Unknown
DEVELOPMENTAL STAGE: Unknown
CELL TYPE: Whole kidney
IMMEDIATE SOURCE:
LIBRARY: Bovine lens cDNA
CLONE: KMC15
POSITION IN GENOME: No. 5424408 known
CHROMOSOME/SEGMENT: No. 5424408 known
PUBLICATION INFORMATION: No. 5424408e
US-07-621-091G-1

Alignment Scores:
Pred. No.: 9,63e-15 Length: 1416
Score: 402.50 Matches: 103
Percent Similarity: 46.62% Conservative: 21
Best Local Similarity: 38.72% Mismatches: 103
Query Match: 15.97% Indels: 40
Gaps: 8

US-09-763-712A-2_COPY_91_547 (1-457) x US-07-621-091G-1 (1-1416)

QY 157 GlnGlyProGlyProArgGlyProArgGlyProArgGlyProArgGlyProGly 176
Db 112 AAAGGAGATCGAGGTCCACCTGGCTCCAGAGGAACCCAGGCGATCGCTCTCTCTGA 171
QY 177 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 191
Db 172 CCTCAGGGAGTCTGTAGAGGCATTAAGAGGACAAAGAGGAGTGTATGGAGGAGCTGCG 231
QY 192 ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArg 211
Db 232 CAAAGAGTCCACCTGGAGCTATAGACATGCGGGTCCACAGGTCATCCGGGAGCACCA 291
QY 212 GlnGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 231
Db 292 GGTGTCCCGGTGACCCAGGCGCCAGAGGTGATCTGATGATTTCCAGGCGATG 351
QY 232 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 251
Db 352 AAAGGGAAGAGGGTAATTTCAGGATTTCCAGGACCATCTGGACCTCCAGGGCAAAAGTGA 411
QY 252 -----LeuPro 253
Db 412 CAAAAGGACCACTGGAGTAGCTGGAGAGCGCTGCACAGTGAAGATCATCTCCCTTCCA 471

QY 254 GlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 273
Db 472 GGAAGCCAGGCCACCTGGTTTCAGCTGGAGAACCCAGGCGATCGAAGGAGAACCCGGGCC 531
QY 274 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 293
Db 532 CCAGGACCCAGGAGATCCAGGACCTGTGGGCAAAAGGTAACCCAGGCGGAGGTGT 591
QY 294 ProGlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluPro 313
Db 592 CCAGCAGGACTCTGGACCACTGGAGAAAAGCAACAAAGGTTGTAAGAGAGAGCAA 651
QY 314 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 329
Db 652 GGACCA--CCTGGATCCGATCCGCTGCCAGGCTTGAAGGGGAAAC--TGGAGACACTGG 707
QY 330 ThrAspLysCysTyrrPheSerValGluLysGluIlePheGluAspAlaLysLeuPhe 349
Db 708 ACC---ACCTGCAGCAGGGGCGAGTATGAGGGCTTGTCTTTACCCGCGACAG----- 758
QY 350 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIle 369
Db 759 ---CCAGACCCAGCAATTCCTCTCTCCAGAGGGAGAGCCGCTCTATAGTGGGT 815
QY 370 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 387
Db 816 TTCTCTCTCTTTGTACAGGAATGAACAAGCCCATGACAGGA----- 860
QY 388 ArgGluAsnGluTrpLys 393
Db 861 CCTGGGAACACTTGGCAG 878

RESULT 12
US-08-399-889-1
; Sequence 1, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Calf
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
US-08-399-889-1

Alignment Scores:
Pred. No.: 9,63e-15 Length: 1416
Score: 402.50 Matches: 103
Percent Similarity: 46.62% Conservative: 21
Best Local Similarity: 38.72% Mismatches: 103
Query Match: 15.97% Indels: 40
Gaps: 8

US-09-763-712A-2_COPY_91_547 (1-457) x US-08-399-889-1 (1-1416)

QY 157 GlnGlyProGlyProArgGlyProArgGlyProArgGlyProArgGlyProGly 176
Db 112 AAAGGAGATCGAGGTCCACCTGGCTCCAGAGGAACCCAGGCGATCGCTCTCTCTGA 171
QY 177 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 191
Db 172 CCTCAGGGAGTCTGTAGAGGCATTAAGAGGACAAAGAGGAGTGTATGGAGGAGCTGCG 231
QY 192 ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProGlyGluArg 211
Db 232 CAAAGAGTCCACCTGGAGCTATAGACATGCGGGTCCACAGGTCATCCGGGAGCACCA 291
QY 212 GlnGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 231
Db 292 GGTGTCCCGGTGACCCAGGCGCCAGAGGTGATCTGATGATTTCCAGGCGATG 351
QY 232 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlu 251
Db 352 AAAGGGAAGAGGGTAATTTCAGGATTTCCAGGACCATCTGGACCTCCAGGGCAAAAGTGA 411
QY 252 -----LeuPro 253
Db 412 CAAAAGGACCACTGGAGTAGCTGGAGAGCGCTGCACAGTGAAGATCATCTCCCTTCCA 471

```
Db 172 CCTCCAGGAGTCCTCTAGAACGCATATAAAGGAGACAAAGGGGTTGATGGAGAGCCTGGC 231
QY 192 ProProGlyProAlaGlyLeuArgGlyProIleGlyProAlaGlyProGlyGluArg 211
Db 232 CAAAGAGGTCCACCTGGAGCTATAGGAGATAGGGGTCCACAGGTATCCGGGAGCACCA 291
QY 212 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 231
Db 292 GGTGTCCCGGTCAGCCAGGAGCCAGAGGTATCTGGATCTATGATTTCAGGATG 351
QY 232 ProGlyProGlnGlyProSerGlyAspGlyProGlyProGlyProGlyLysGly 251
Db 352 AAAGGAAGAGGTAATTCAGATTTCAGGACCACTGGACCACTCCAGGCAAGAGTGA 411
QY 252 -----LeuPro 253
Db 412 CCAAAAGGACCACCTGGAGTACGTGGAGAGCCTGGACAGTGAAGATCATCTCCCTTCCA 471
QY 254 GlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 273
Db 472 GGAAGCCCAAGGCCACCTGGTTCAGCTGGAGAACCCAGGATGCAAGGAGAACCCGGGCC 531
QY 274 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 293
Db 532 CCAGGACCACAGGAGATCCAGGACCTGTGGCCAAAAGGTAAACAGGAGGATGTT 591
QY 294 ProProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 313
Db 592 CCACCAAGAACTCCTGGACCACTGGAGAAAAAGGCAACAAAGTTGTAAGAGAGACAA 651
QY 314 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 329
Db 652 GGACCA---CCTGGATCCAGTCCGCTGGAGGCTTGAAGGGGAAACC-TGGAGACACTGG 707
QY 330 ThrAspLysCysTyrTyrPheSerValGluLysGluLysGluLysPheLeuPhe 349
Db 708 ACC---ACCTGACAGAGGGGAGTATGAGGGCTTTGCTTTACCCGGCACAG-----758
QY 350 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIle 369
Db 759 ---CCAGACCACAGCAATTCCTCTGTCCAGAGGACAGAGCCGCTCTATAGTGGTT 815
QY 370 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 387
Db 816 TTCTCTTCTTTGTACAGGAAATGAACAAGCCCATGGACAGCA-----860
QY 388 ArgGluAsnGluTrpLys 393
Db 861 CCTGGGAACACTTGGCAG 878

RESULT 13
US-09-167-364-1
; Sequence 1, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Calf
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
```

```
US-09-167-364-1
Alignment Scores:
Pred. No.: 9,63e-15 Length: 1416
Score: 402.50 Matches: 103
Percent Similarity: 46.62% Conservative: 21
Best Local Similarity: 38.72% Mismatches: 103
Query Match: 15.97% Indels: 40
DB: 3 Gaps: 8

US-09-763-712a-2_COPY_91_547 (1-457) x US-09-167-364-1 (1-1416)
QY 157 GlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGly 176
Db 112 AAAGGAGATCGAGGTCCACCTGGCTCCAGAGGAAACAGGATGCTGCTCTCTGGA 171
QY 177 ProThrGlyAsn-----LysGlyGlnLysGlyGluLysGlyGluProGly 191
Db 172 CCTCCAGGAGTCCCTGTAGAACGCATATAAAGGAGACAAAGGGTGTGATGGAGAGCCTGGC 231
QY 192 ProProGlyProAlaGlyLeuArgGlyProIleGlyProAlaGlyProProGlyGluArg 211
Db 232 CAAAGAGGTCCACCTGGAGCTATAGGAGATAGGGGTCCACAGGTATCCGGGAGCACCA 291
QY 212 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 231
Db 292 GGTGTCCCGGTCAGCCAGGAGCCAGAGGTATCTGGATCTATGATTTCAGGATG 351
QY 232 ProGlyProGlnGlyProSerGlyAspGlyProGlyProGlyProGlyLysGly 251
Db 352 AAAGGAAGAGGTAATTCAGATTTCAGGACCACTGGACCACTCCAGGCAAGAGTGA 411
QY 252 -----LeuPro 253
Db 412 CCAAAAGGACCACCTGGAGTACGTGGAGAGCCTGGACAGTGAAGATCATCTCCCTTCCA 471
QY 254 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 273
Db 472 GGAAGCCCAAGGCCACCTGGTTCAGCTGGAGAACCCAGGATGCAAGGAGAACCCGGGCC 531
QY 274 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 293
Db 532 CCAGGACCACAGGAGATCCAGGACCTGTGGCCAAAAGGTAAACAGGAGGATGTT 591
QY 294 ProProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 313
Db 592 CCACCAAGAACTCCTGGACCACTGGAGAAAAAGGCAACAAAGTTGTAAGAGAGACAA 651
QY 314 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 329
Db 652 GGACCA---CCTGGATCCAGTCCGCTGGAGGCTTGAAGGGGAAACC-TGGAGACACTGG 707
QY 330 ThrAspLysCysTyrTyrPheSerValGluLysGluLysGluLysPheLeuPhe 349
Db 708 ACC---ACCTGACAGAGGGGAGTATGAGGGCTTTGCTTTACCCGGCACAG-----758
QY 350 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIle 369
Db 759 ---CCAGACCACAGCAATTCCTCTGTCCAGAGGACAGAGCCGCTCTATAGTGGTT 815
QY 370 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 387
Db 816 TTCTCTTCTTTGTACAGGAAATGAACAAGCCCATGGACAGCA-----860
QY 388 ArgGluAsnGluTrpLys 393
Db 861 CCTGGGAACACTTGGCAG 878

RESULT 14
US-09-439-897-1
; Sequence 1, Application US/09439897
; Patent No. 6277558
; GENERAL INFORMATION:
```

```
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1416)
US-09-439-897-1

Alignment Scores:
Pred. No.: 9,638-15 Length: 1416
Score: 402.50 Matches: 103
Percent Similarity: 46.62% Conservatives: 21
Best Local Similarity: 38.72% Mismatches: 103
Query Match: 15.97% Indels: 40
DB: 4 Gaps: 8

US-09-763-712A-2_COPY_91_547 (1-457) x US-09-439-897-1 (1-1416)
QY 157 GlnGlyProGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGly 176
Db 112 AAAGGAGATCGAGGTCCACCTGGCTCCAGAGAAACCCAGGCATGCTGCTCTCTGGA 171
QY 177 ProThrGlyAsn-----LysGlyGlnLysGlyGlyGlyGluProGly 191
Db 172 CCTCAGGAGGATCCTGTAGAAAGCATAAAGAGACAAAGGGTTGATGGAGAGCCTGGC 231
QY 192 ProGlyProAlaGlyGluArgGlyProGlyProAlaGlyProGlyGluArg 211
Db 232 CAAAGAGGTCCACCTGGGCTTAGGACATACAGGTCATCCGGGAGACCA 291
QY 212 GlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLys 231
Db 292 GGTGTCCCGGTCCAGCCAGGGCCAGAGGTGATCTGGATTCTATGATTTCAGGCATG 351
QY 232 ProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyLysGluGly 251
Db 352 AAAGGGAAGAGGTAATTCAGGATTTCCAGGACCACTTGGACCTCCAGGGCAAAAGTGA 411
QY 252 -----LeuPro 253
Db 412 CCAAAAGGACCACCTGGAGTACGTGGAGAGCCCTGGCACAGTGAAGATCATCTCCCTTCCA 471
QY 254 GlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 273
Db 472 GGAAGCCAGGCCACCTGTTCCAGCTGGAGAACACAGGATGCAAGAGAACCCGGGCC 531
QY 274 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 293
Db 532 CCAGGACCACAGGAGATCCAGGACCTGTGGGCCAAAGGTAACACAGGGAGGATGTT 591
QY 294 ProGlyProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 313
Db 592 CCACAGGAACTCCTGGACCACTGGAGAAAGGCAACAAAGGTTGTAAAGAGAGCA 651
QY 314 ThrProAlaProGluAspAsnGlyCysPro-----ProHisTrpLysAsnPhe 329
Db 652 GGACCA--CCTGGATCCCATGGCTGCGCCAGGCTTCAAGGGGAAACC-TGGAGACACTGG 707
QY 330 ThrAspLysCysTyrTrpPheSerValGlyLysGluPheGluAspAlaLysLeuPhe 349
Db 708 ACC---ACCTGCAGCAGGCGCATGATGAGGGCTTTGTCTTTATCCCGGCACAG----- 758
QY 350 CysGluAspLysSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIle 369
Db 759 ----CCAGACCACAGCAATTCCTCTCTGTCAGAAAGGAGGACAGACCCGCTCTATAGTGGTT 815
```

```
QY 370 LysLysGlnMetValGlyArgGlu-----SerHisTrpIleGlyLeuThrAspSerGlu 387
Db 816 TTCTCTCTCTCTTTGTACAAAGGAATGAACAAGCCCATGGACAGGA-----860
....QY 388 ArgGluAsnGluTrpLys 393
Db 861 CCTGGGAACACTTGGCAG 878

RESULT 15
US-09-484-970B-4
; Sequence 4, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 4359
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 331365 (1553795CB1)
US-09-484-970B-4

Alignment Scores:
Pred. No.: 3,998-14 Length: 4359
Score: 400.50 Matches: 90
Percent Similarity: 54.89% Conservatives: 11
Best Local Similarity: 48.91% Mismatches: 64
Query Match: 15.89% Indels: 19
DB: 4 Gaps: 7

US-09-763-712A-2_COPY_91_547 (1-457) x US-09-484-970B-4 (1-4359)
```

```
QY 156 LeuGlnGlyProGlyProArgGlyProArgGlyAspArg-----169
Db 448 CTGCCAGGACCCCGAGGCCCGCCAGGTGCCAAAGGGCTCCGAGGAATCCAGGCTTCGCA 507
QY 170 GlySerGlnGlyProGlyProThrGlyProThrGlyAsnLysGlyGlnLysGlyGlu 189
Db 508 GGAGCTGATGGAGGACGAGGTCCAGGGCTTCCAGGAGACGCGAGTCTGTAAGGTTTC 567
QY 190 ProGlyProPro-----GlyProAlaGlyGluArgGlyProIleGlyProAlaGly 206
Db 568 CCAGGACCCCGAGGTTTCATAGGACCCCGAGGATCCAAAGTCCAGTGGGCTCCCTGGC 627
QY 207 ProGlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArg 226
Db 628 CCAGATGGATCCCGAGGTCCATCGGCTGCCAGGCCAGATGGGCCCCCTGGGGAAGG 687
QY 227 GlySerProGlyLysPro---GlyProGln---GlyProSerGlyAspProGlyProPro 244
Db 688 GGCTCCCTGGAGAGATCTCTGGGAGTCCAGCCCGGCGGAGATGCTGTGTGCTCT 747
QY 245 GlyProGlyLysGlyGluGlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeu 264
Db 748 GGACAGCTTGGCTTAAAGGCTTCCCGAGACAGAGGCCCCCTGGATTTCAGAGNAGC 807
QY 265 GlnGlyThrValGlyGlyProGlyValProGlyProArgGlyLeuProGlyLeuProGly 284
Db 808 CAAGGATG-----CCTGGGATGCCAGGGCTGAAGGGCCAGCCAGGCTCCAGGA 858
QY 285 ValProGlyMetProGlyProLysGlyProGlyPro-----ProGlyProSer 301
Db 859 CCTTCCGCGCAGCCAGGCTGTATGGGCTTCCAGGACTGCATGATGATTCAGGAGCTCCT 918
```

QY 302 GlyAlaValProLeuAlaLeuGlnAsnGluPro-----ThrProAlaProGluAsp 319
Db 919 GGCCAAAGAGGGCCCTTGGGGCTGCCAGGAATCCAGGGCGTGAAGGTCTGCCTGGTGAT 978
QY 320 AsnGlyCysPro 323
Db 979 AGAGGGGACCCCT 990

Search completed: March 21, 2003, 12:34:17
Job time : 61.552 secs

GenCore version 5.1.4.p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:23:12 ; Search time 222.049 Seconds
(without alignments)
1596.010 Million cell updates/sec

Title: US-09-763-712A-2_COPY_91_547

Perfect score: 2520

Sequence: 1 MNSOLNFTGOMENTTISQ.....EDVNNFICKDRETVLSSAL 457

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09763712/runat_14032003.100950.18119/app_query.fasta.1.1877
-DB=PublishedApplications_NA -QFM=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPECL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Diosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09763712@cgn_1.1.123 @runat_14032003.100950.18119
-NCPU=6 -ICPU=3 -NO_XLPPY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications_NA:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------------------------|
| 1 | 2514 | 99.8 | 2930 | 10 | US-09-745-763-198 Sequence 198, App |
| 2 | 474.5 | 18.8 | 1410 | 10 | US-09-954-456-2274 Sequence 2274, App |
| 3 | 455 | 18.1 | 1133 | 9 | US-09-924-340-57 Sequence 57, Appl |
| 4 | 455 | 18.1 | 1133 | 9 | US-09-992-600A-57 Sequence 57, Appl |

| | | | | | |
|----|-------|------|------|----|--------------------------------------|
| 5 | 453 | 18.0 | 6158 | 10 | US-09-919-497-6 Sequence 6, Appl |
| 6 | 453 | 18.0 | 6158 | 10 | US-09-954-456-762 Sequence 762, App |
| 7 | 452.5 | 18.0 | 6728 | 10 | US-09-954-456-782 Sequence 782, App |
| 8 | 452.5 | 18.0 | 6728 | 10 | US-09-880-107-3946 Sequence 3946, Ap |
| 9 | 444 | 17.6 | 5416 | 10 | US-09-954-456-786 Sequence 786, App |
| 10 | 444 | 17.6 | 5416 | 10 | US-09-880-107-2094 Sequence 2094, Ap |
| 11 | 440.5 | 17.5 | 4149 | 12 | US-10-044-090-104 Sequence 104, App |
| 12 | 440.5 | 17.5 | 5086 | 10 | US-09-880-107-3947 Sequence 3947, Ap |
| 13 | 440.5 | 17.5 | 5145 | 9 | US-09-925-299-206 Sequence 206, App |
| 14 | 440.5 | 17.5 | 5145 | 10 | US-09-925-299-206 Sequence 206, App |
| 15 | 440.5 | 17.5 | 5432 | 12 | US-10-044-090-22 Sequence 22, Appl |
| 16 | 438 | 17.4 | 2235 | 10 | US-09-962-436-308 Sequence 308, App |
| 17 | 435 | 17.3 | 3690 | 12 | US-10-044-090-448 Sequence 448, App |
| 18 | 433 | 17.2 | 2192 | 10 | US-09-925-301-42 Sequence 42, Appl |
| 19 | 432.5 | 17.2 | 1486 | 10 | US-09-925-302-247 Sequence 247, App |
| 20 | 431.5 | 17.1 | 3226 | 10 | US-09-954-456-725 Sequence 725, App |
| 21 | 426.5 | 16.9 | 4908 | 9 | US-10-001-887-33 Sequence 33, Appl |
| 22 | 425.5 | 16.9 | 2520 | 10 | US-09-880-107-3685 Sequence 3685, Ap |
| 23 | 423 | 16.8 | 481 | 10 | US-09-815-343-1516 Sequence 1516, Ap |
| 24 | 418.5 | 16.6 | 511 | 9 | US-09-736-457-1138 Sequence 1138, Ap |
| 25 | 418.5 | 16.6 | 511 | 9 | US-09-902-941-1138 Sequence 1138, Ap |
| 26 | 418.5 | 16.6 | 511 | 9 | US-09-849-626-1138 Sequence 1138, Ap |
| 27 | 415.5 | 16.5 | 2823 | 10 | US-09-919-497-7 Sequence 7, Appl |
| 28 | 414.5 | 16.4 | 2542 | 9 | US-09-954-531-961 Sequence 961, App |
| 29 | 414.5 | 16.4 | 2542 | 10 | US-09-964-824A-255 Sequence 255, App |
| 30 | 409 | 16.2 | 3394 | 10 | US-09-880-107-2178 Sequence 2178, Ap |
| 31 | 407 | 16.2 | 477 | 10 | US-09-815-343-1517 Sequence 1517, Ap |
| 32 | 397.5 | 15.8 | 2691 | 10 | US-09-925-302-64 Sequence 64, Appl |
| 33 | 394.5 | 15.7 | 1797 | 9 | US-09-978-285A-613 Sequence 613, App |
| 34 | 394.5 | 15.7 | 1797 | 9 | US-09-978-697-613 Sequence 613, App |
| 35 | 394.5 | 15.7 | 1797 | 9 | US-09-999-832A-613 Sequence 613, App |
| 36 | 394.5 | 15.7 | 1797 | 9 | US-09-978-189-613 Sequence 613, App |
| 37 | 394.5 | 15.7 | 1797 | 9 | US-10-174-590-331 Sequence 331, App |
| 38 | 394.5 | 15.7 | 1797 | 9 | US-10-176-758-331 Sequence 331, App |
| 39 | 394.5 | 15.7 | 1797 | 9 | US-10-175-737-331 Sequence 331, App |
| 40 | 394.5 | 15.7 | 1797 | 9 | US-10-175-737-331 Sequence 331, App |
| 41 | 394.5 | 15.7 | 1797 | 9 | US-10-175-738-331 Sequence 331, App |
| 42 | 394.5 | 15.7 | 1797 | 9 | US-10-175-752-331 Sequence 331, App |
| 43 | 394.5 | 15.7 | 1797 | 9 | US-10-176-482-331 Sequence 331, App |
| 44 | 394.5 | 15.7 | 1797 | 9 | US-10-176-482-331 Sequence 331, App |
| 45 | 394.5 | 15.7 | 1797 | 9 | US-10-176-757-331 Sequence 331, App |

ALIGNMENTS

RESULT 1
US-09-745-763-198
; Sequence 198, Application US/09745763
; Patent No. US20020085394A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John M.
; LaVallie, Edward R.
; Collins-Racie, Lisa A.
; Evans, Cheryl
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; ENCODING THEM
; NUMBER OF SEQUENCES: 219
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/745,763
 ; FILING DATE: 18-Jun-2000
 ; CLASSIFICATION: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Sprunger, Suzanne A.
 ; REGISTRATION NUMBER: 41,323
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 498-8284
 ; TELEFAX: (617) 876-5851
 ; INFORMATION FOR SEQ ID NO: 198:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2930 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 198:
 ; US-09-745-763-198

Alignment Scores:
 Pred. No.: 4,13e-130 Length: 2930
 Score: 2514.00 Matches: 456
 Percent Similarity: 99.78% Conservative: 0
 Best Local Similarity: 99.78% Mismatches: 1
 Query Match: 99.76% Indels: 0
 DB: 10 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x US-09-745-763-198 (1-2930)

QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
 DB 921 ATGAACAGCCAGCTCAACTCAATTCACAGGTGAGGAGAACATCACCACTATCTCTCAA 980
 QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnThr 40
 DB 981 GCCACGACGACAGCTGAAGACCTGCAGGACTTACACAAAGATGCAGAGAAATAGAAC 1040
 QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
 DB 1041 GCCATCAAGTTCAACCACTGAGGAGAACGCTTCCAGCTCTTTGAGACGATATTGGAAC 1100
 QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
 DB 1101 ATCATAGCAATATCAGTTACACAGCCACCACTCGGACGCTGACCACTCAATTAAT 1160
 QY 81 GluValArgThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
 DB 1161 GAAGTCAGGACCACTTGCACATACCTTACCAACACACAGATGATCTGACCTCTTGT 1220
 QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
 DB 1221 AATAATACCTGGCCAAACATCGTTGGATTCTCTCTCAGGATGCAACAAGATTGTG 1280
 QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetLys 140
 DB 1281 ATGAGTCGAGTTGACACTGAAGTAGCAACTATATCAGTATTATGGAAGAAATGAAG 1340
 QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
 DB 1341 CTAGTAGACTCCAGCATGGTCAGTCATCAAGAAATTTTACAACTACTACAGGTCCACG 1400
 QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProThrGlyAsn 180
 DB 1401 GGCCCCAGGGTCCAAGAGGTGACAGAGGATCCAGGGGACCCCTGGCCCACTGGCAAC 1460
 QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGlyGluArg 200
 DB 1461 AAGGGACAGAAAGAGAGAGAGGGGAGCGCTGGACCACTGGCCCTCGGGGTGAGAGAGC 1520
 QY 201 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 DB 1521 CCATTTGGACAGCTGGTCCCCCGGAGAGCGTGGCGGCAAGGATCTAAGGCTCCACG 1580

QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
 DB 1581 GGCCCCAAAGGCTCCCGTGGTTCCCTCGGAAGCCCGGCCCTCAGGCCCCAGTGGGAC 1640
 QY 241 ProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGlnGlyProGly 260
 DB 1641 CCAGGCCCGCCCGGCCACACAGGCAAGAGGACTCCCGGCCCTCAGGCCCCCTCTGTC 1700
 QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 DB 1701 TTCAGGGACTTCAGGGCACCTTGGGAGCGCTGGGGTGGCTGGACCTCGGGAGTGC 1760
 QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
 DB 1761 GGCTTGGCTGGGTACACAGGATGCCAGGCCCAAGGGCCCCCGGCCCTCTGGCCCA 1820
 QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 DB 1821 TCAGGAGCGGTGGTGGCTGGCCCTGCAGAAATGAGCCAACCCCGCACGGAGGACAAT 1880
 QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
 DB 1881 AGCTGCCCGCCCTACTGGGAAGAACTTCACAGACAATGCTACTATTTTCAGTTGAGAAA 1940
 QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
 DB 1941 GAAATTTTGGAGATGCAAGCTTTCTGTGAAGACAAGCTTTCACATCTCTGTTTCATA 2000
 QY 361 AsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 DB 2001 AACACTAGAGAGAACAGCAATGGATAAAAAAACAGATGGTAGGAGAGAGAGACCACTGG 2060
 QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTyrLysTrpLeuAspGlyThrSerPro 400
 DB 2061 ATCGCCCTCACAGCTCAGAGCGTGAAATGAATGAAGTGGCTGGATGGGACATCTCCA 2120
 QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
 DB 2121 GACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCATGGCCATGGCCAGCA 2180
 QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 DB 2181 GAAGACTGTGCTGGGTGATTATCTGGGCAGTGGAGGATTTCCAATGTGAAGACGTC 2240
 QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 DB 2241 AATACTTCAATTTGGAAAAAGACAGGAGGACAGTACTGTCTATCTGCATTA 2291

RESULT 2

; US-09-954-456-2274
 ; Sequence 2274, Application US/09954456
 ; Patent No. US20020115057A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Young, Paul
 ; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
 ; TITLE OF INVENTION: Sets
 ; FILE REFERENCE: 689290-76
 ; CURRENT APPLICATION NUMBER: US/09/954,456
 ; CURRENT FILING DATE: 2001-09-18
 ; PRIOR APPLICATION NUMBER: US/60/233,617
 ; PRIOR FILING DATE: 2000-09-18
 ; PRIOR APPLICATION NUMBER: US/60/234,052
 ; PRIOR FILING DATE: 2000-09-20
 ; PRIOR APPLICATION NUMBER: US/60/234,923
 ; PRIOR FILING DATE: 2000-09-25
 ; PRIOR APPLICATION NUMBER: US/60/235,134
 ; PRIOR FILING DATE: 2000-09-25
 ; PRIOR APPLICATION NUMBER: US/60/235,637
 ; PRIOR FILING DATE: 2000-09-26
 ; PRIOR APPLICATION NUMBER: US/60/235,638
 ; PRIOR FILING DATE: 2000-09-26
 ; PRIOR APPLICATION NUMBER: US/60/235,711

PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,720
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,840
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,863
PRIOR FILING DATE: 2000-09-27
NUMBER OF SEQ ID NOS: 2276
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2274
LENGTH: 1410
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-456-2274

Alignment Scores:
Pred. No.: 1.68e-18 Length: 1410
Score: 474.50 Matches: 111
Percent Similarity: 49.57% Conservative: 41
Best Local Similarity: 36.27% Mismatches: 132
Query Match: 18.83% Indels: 23
DB: 6 Gaps: 6

US-09-763-712A-2_COPY_91_547 (1-457) x US-09-954-456-2274 (1-1410)

| | | | |
|----|------|---|------|
| Qy | 156 | LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro | 175 |
| Db | 400 | ATGCCTGGCAAGCTGGCCCATGTGGGCCAAAGGGACAATGGCTCTGTGGAGAACCT | 459 |
| Qy | 176 | GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyPro | 195 |
| Db | 460 | GGACCAAAGGAGACACTGGCCCAAGTGGACCTCCAGGACCTCCCGGTGTCCTGGTGCCA | 519 |
| Qy | 196 | AlaGlyGluArGlyProIleGlyProAlaGlyProGlyGluArGlyGlyLysGly | 215 |
| Db | 520 | GCTGAAGAGAGTCCCTGGGAAGCAGGGGAACATAGGACCTCAGGGCAAGCCAGGC | 579 |
| Qy | 216 | SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln | 235 |
| Db | 580 | CCTGAAGAGAGTGGCCCAAGAGAGTAAGTGGCCCAAGGACATGAGGCTCGGCA | 639 |
| Qy | 236 | GlyProSerGlyAspProGlyProProGlyPro-----ProGlyLysGluGlyLeu | 252 |
| Db | 640 | GGGCAAGAGGCCTCGAGCCCTAAGGGAGAGGAGGTGTCCTGTGGTGGAGTGC | 699 |
| Qy | 253 | ProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGly | 272 |
| Db | 700 | CCTGAAGAGGAGGAGGAGGAGGAGTCTCTGGAGCATGGTCCCAGGGAAGTCCAGGT | 759 |
| Qy | 273 | ValProGlyProArGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys | 292 |
| Db | 760 | GCCAGGGAGCCCCGGGATTGAAGGGGACAAAGGCATCTCTGGAGACAAAGGACAAAG | 819 |
| Qy | 293 | GlyProProGlyProProGlyProSerGlyAlaValProLeu---AlaLeuGlnAsn | 311 |
| Db | 820 | GGAGAAAGTGGGTTCCAGATGTTGCTCTGAGGCAGCAGGTGTGAGCCTTACAGGA | 879 |
| Qy | 312 | GluProThrProAlaProGluAsnGlyCys-----GGCTGCTTTCTCTCAGTATAAGAAGTTGAGCT | 929 |
| Db | 880 | CAAGT-ACAGACCTCCA-----GGCTGCTTTCTCTCAGTATAAGAAGTTGAGCT | 929 |
| Qy | 323 | ProHisTrp-LysAsnPhetHrAspLysCystyrTyrrPheSerValGluLysGluIl | 342 |
| Db | 930 | CTTCCCAATGCCAAGTGTGGGAGAGATTTTCAAGACAGCAGGCTTTGTAACC | 989 |
| Qy | 342 | ePheLuAspaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnTh | 362 |
| Db | 990 | ATTACGGAGGACAGCTGCTGTGCACACAGCGCTGGTGCACAGTGTGGCCTCTCCACGCTC | 1049 |
| Qy | 362 | rArgGluGlnGlnTrpIleLysLysGlnMetValGlyarg---GluSerHisTrpIl | 381 |
| Db | 1050 | TGCCGCTGAGAATGCCGCTTGTGAACAGCTGCTGTAGTAAGAACAGGAGTGTCTTCT | 1109 |

```
Db 260 CTGGATTCCAGGAAAGTTGGATCACCTGGCCACCTGGCCCTCAAGCAGAGAAGGC 319
Qy 216 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyPro 234
Db 320 AGCGAAGGATTGCGAGGCCATCAGGCTGCTGGCTCCCTGGGCCACCGGACCTCCT 379
Qy 235 -----GlnGlyProSer 238
Db 380 GGGATTTCAGGGCCCCCGCTGGATGTTGGATGGGAAGATGGCAAGCCTGGCTTG 439
Qy 239 ---GlyAspProGlyProGlyProPro----- 247
Db 440 AGGGGGGACCTTGGTCTGCTGGCCCTGGACTCATGGACCACCGGGCTTTAAGGG 499
Qy 248 -----GlyLysGluGlyLeuProGln----- 259
Db 500 AAACAGGACATCCTGGCTCCAGAGGCTAAGGGTGACTGGGCAACACGAGTCTCCT 559
Qy 260 GlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeu 279
Db 560 GCGACACTGGCGGCTGGCGCAGAGGTTGAACCTGGTGCCATGGGACCCCGAGGAAGA 619
Qy 280 ProGlyLeuPro----- 283
Db 620 CCGGTCCCCCGGACACCTGGGCCACGAGGCCCTCCAGGCCAGCCAGCAGCTGGG 679
Qy 284 ----- 284
Db 680 ATCTCTGCACTGGTCTGAAAGAGACCAGGAGGCCACCGAGAAAGGGGCCCTTGCAGGC 739
Qy 285 ValProGlyMetProGlyProLysGlyProGlyProPro-----GlyProSerGly 302
Db 740 CTCCCGAGCCAGCCCGGCCCTCCAGAGGTTGAACCTGGTGCCATGGGACCCCGAGGAGTGGGT 799
Qy 303 AlaValValProLeuAlaLeuGlnAsnGluPro---ThrProAlaPro 317
Db 800 GCAACAGGACCAATGGGCCAGCAAGGCATCCCTGGCATCCCTGGGCC 847

RESULT 4
US-09-992-600A-57
; Sequence 57, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjanin, Stephanie
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US4.DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: JPatent
; SEQ ID NO 57
; LENGTH: 1133
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..475
; NAME/KEY: CDS
```

```
; LOCATION: 476..964
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 965..1133
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1101..1106
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 1118..1133
US-09-992-600A-57

Alignment Scores:
Pred. No.: 1,62e-17 Length: 1133
Score: 455.00 Matches: 98
Percent Similarity: 47.46% Conservative: 14
Best Local Similarity: 41.53% Mismatches: 50
Query Match: 18.06% Indels: 74
DB: 9 Gaps: 7

US-09-763-712A-2_COPY_91_547 (1-457) x US-09-992-600A-57 (1-1133)
Qy 156 LeuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 175
Db 140 CTCCAGGCCCCCAGGCCAGCGGTCCAGAGGAGCAGGACCCCAAGTAACCTCC 199
Qy 176 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 195
Db 200 GGTGAGAAGGGGACCCAGGAGTTCAAGGCCAGCCAGGCTTCCGGGCCACCGGGTCCC 259
Qy 196 AlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGly 215
Db 260 CTGGATTCCAGGCAAGTTGGATCACCTGGCCACCTGGCCCTCAAGCAGAGAAGGC 319
Qy 216 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyPro 234
Db 320 AGCGAAGGATTGCGAGGCCATCAGGCTGCTGGCTCCCTGGGCCACCGGACCTCCT 379
Qy 235 -----GlnGlyProSer----- 238
Db 380 GGGATTTCAGGGCCCCCGCTGGATGTTGGATGGGAAGATGGCAAGCCTGGCTTG 439
Qy 239 ---GlyAspProGlyProProPro----- 247
Db 440 AGGGGGGACCTTGGTCTGCTGGCCCTGGACTCATGGACCACCGGGCTTTAAGGG 499
Qy 248 -----GlyLysGluGlyLeuProGlyProGln----- 259
Db 500 AAACAGGACATCCTGGCTCCAGAGGACCTAAGGGTGACTGGGCAACACGAGTCTCCT 559
Qy 260 GlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeu 279
Db 560 GCGACACTGGCGGCTGGCGCAGAGGTTGAACCTGGTGCCATGGGACCCCGAGGAAGA 619
Qy 280 ProGlyLeuPro----- 283
Db 620 CCGGTCCCCCGGACACCTGGGCCACGAGGCCCTCCAGGCCAGCCAGCAGCTGGG 679
Qy 284 ----- 284
Db 680 ATCTCTGCACTGGTCTGAAAGAGACCAGGAGGCCACCGAGAAAGGGGCCCTTGCAGGC 739
Qy 285 ValProGlyMetProGlyProLysGlyProGlyProPro-----GlyProSerGly 302
Db 740 CTCCCGAGCCAGCCCGGCCCTCCAGAGGTTGAACCTGGTGCCATGGGACCCCGAGGAGTGGGT 799
Qy 303 AlaValValProLeuAlaLeuGlnAsnGluPro---ThrProAlaPro 317
Db 800 GCAACAGGACCAATGGGCCAGCAAGGCATCCCTGGCATCCCTGGGCC 847

RESULT 5
US-09-919-497-6
; Sequence 6, Application US/09919497
```

```
; Patent No. US2002010662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 6158
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Unsure
; LOCATION: (2434)..(2434)
; OTHER INFORMATION: n = a, c, g or t/u
US-09-919-497-6

Alignment Scores:
Pred. No.: 9,61e-17 Length: 6158
Score: 453.00 Matches: 130
Percent Similarity: 39.81% Conservative: 38
Best Local Similarity: 30.81% Mismatches: 76
Query Match: 17.98% Indels: 180
DB: 10 Gaps: 18

US-09-763-712A-2_COPY_91_547 (1-457) x US-09-919-497-6 (1-6158)
QY 158 GlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyPro 177
DB 4032 GGACCTCCAGGTGCGCAAGGGCGCCAGGTGATGATGCCCTAAGGTAACCGGTCCT 4091
QY 178 Thr----- 178
DB 4092 GTTGGTTTCTCGAGATCTCGTCTCTGGGAACTTGGCCCTGCAGGTCAAGATGCT 4151
QY 179 -----GlyAsnLysGlnLysGlyGluLysGlyGluProGlyProGlyProAla 196
DB 4152 GTTGGTGTGCAAGGGTGAAGATGAGATCTGTCAACCGGGTCTCTCGGCCCACT 4211
QY 197 GlyGluArgGlyProIleGlyPro-----AlaGlyProProGlyGluArgGly 213
DB 4212 GGTGAGGTGGCCACCAGGTCTCTTGAAACAGGTCTCTCTGGAGCTCCAGGTGCA 4271
QY 214 LysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysPro 232
DB 4272 GAGGGAAGACAAAGTGAAGAGGTGTAAGGGGAAGCAGGTGCAGAAAGGTCTCTCGGA 4331
QY 233 -----GlyProGlnGlyProSerGlyAspProGlyPro----- 243
DB 4332 AAAACCGGCCCACTCGGTCTCAGGACCTCGAGGAAAGCCCTGGTCCAGAGGTCTTCGG 4391
QY 244 -----ProGlyProGlyLysGluGlyLeu----- 252
DB 4392 GGCATCCCTGGTCTGTGGGAACAAGTCTCCCTGGAGCTGCAGGCCAAGATGACCA 4451
QY 253 ProGlyProGlnGlyProProGlyPhe----- 261
DB 4452 CCTGGTCTATGGGACCTCTCGCTTACCTGTCTCAAAGGTGACCTGGCTCCCAAGGT 4511
QY 261 ----- 261
DB 4512 GAAAGGAGACATCTCGGTGTTTAATGGCTGATTGTCCTCCAGGAGAACAGGGGAAAA 4571
QY 262 -----GlnGlyLeuGlnGlyThrValGlyGluPro-----GlyVal 273
DB 4572 GTGACCGAGGCTCTCTGGAACTCAAGGATCTCCAGGACAAAGGGATGGGGAAAT 4631
QY 274 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys 292
DB 4632 CCTGTGCTGTGGTCCCTTAGGTCCACCTGTGCTCCTCCAGGCTTACCAGGTCTCTCAAGGC 4691
QY 292 ----- 292
DB 4692 CCAAGGGTAACAAGGCTCTACTGGACCCGCTGCCCAAGAAAGGTGCACAGTGGTCTTCCA 4751
QY 293 GlyProProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGlu 312
DB 4752 GGGCTCTCTGGGCTCCAGGTCCAGGTCCAGTGTGAAGTCATT-----CAG 4793
QY 313 ProThrPro-AlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAsp 332
DB 4794 CTTTACCAATCT-----TGCTCTCCAAAAAACA----- 4824
QY 332 sCystyrtyrPheSerValGluLysGluPheGluAspAlaLysLeuPheCysGluAs 352
DB 4825 -----GAAGACATCTCAAGGCATGCAAG-- 4848
QY 352 pLysSerSerHisLeuValPheIleAsnThrArgGluGluGln----- 366
DB 4849 -CAGATGCAGATGATAATATTCTTGATTACTCGGATGGAATGAAGAAATATTTGGTTCC 4907
QY 367 -----GlnTrpIleLysLysGlnMe 373
DB 4908 CTCATATTCCTGAAACAGACATCGAGCATATGAAATTTCCAAATGGTACTCAGACCAAT 4967
QY 373 tValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrp 393
DB 4968 CCA-GCCCGAAGTCTTAAGACCTCAACTGACCATCTGACTTCCCAGATGGTGAATA 5026
QY 393 strLeuAsp-----GlyThrSerProAsp-----TyrLysAsnTrp 406
DB 5027 TTGGATTGATCTCAACCAAGGTGTCTCAGGAGATCTCTTCAAAGTTTACTGTGAATTCAC 5086
QY 406 sAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGly 426
DB 5087 ATCTGGT-----GGTGAGACTTGC----- 5105
QY 426 uIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAspPheLysG 446
DB 5106 -ATTATTCGA-----GACAAAAATCTGAGGGAGTAAG-AAATTCATCATGCGC 5151
QY 446 uLys 447
DB 5152 CAAA 5155
RESULT 6
US-09-954-456-762
; Sequence 762, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
```


| | |
|--|---|
| APPLICANT: | Young, Paul |
| TITLE OF INVENTION: | Process for Identifying Anti-Cancer Therapeutic Agents Using Cand |
| FILE REFERENCE: | 689290-76 |
| CURRENT APPLICATION NUMBER: | US/09/954,456 |
| CURRENT FILING DATE: | 2001-09-18 |
| PRIOR APPLICATION NUMBER: | US/60/233,617 |
| PRIOR FILING DATE: | 2000-09-18 |
| PRIOR APPLICATION NUMBER: | US/60/234,052 |
| PRIOR FILING DATE: | 2000-09-20 |
| PRIOR APPLICATION NUMBER: | US/60/234,923 |
| PRIOR FILING DATE: | 2000-09-25 |
| PRIOR APPLICATION NUMBER: | US/60/235,134 |
| PRIOR FILING DATE: | 2000-09-25 |
| PRIOR APPLICATION NUMBER: | US/60/235,637 |
| PRIOR FILING DATE: | 2000-09-26 |
| PRIOR APPLICATION NUMBER: | US/60/235,638 |
| PRIOR FILING DATE: | 2000-09-26 |
| PRIOR APPLICATION NUMBER: | US/60/235,711 |
| PRIOR FILING DATE: | 2000-09-27 |
| PRIOR APPLICATION NUMBER: | US/60/235,720 |
| PRIOR FILING DATE: | 2000-09-27 |
| PRIOR APPLICATION NUMBER: | US/60/235,840 |
| PRIOR FILING DATE: | 2000-09-27 |
| PRIOR APPLICATION NUMBER: | US/60/235,863 |
| PRIOR FILING DATE: | 2000-09-27 |
| NUMBER OF SEQ ID NOS: | 2276 |
| SOFTWARE: | PatentIn version 3.0 |
| SEQ ID NO | 786 |
| LENGTH: | 5416 |
| TYPE: | DNA |
| ORGANISM: | Homo sapiens |
| US-09-954-456-786 | |
| Alignment Scores: | |
| Pred. No.: | 2,67e-16 |
| Score: | 444.00 |
| Percent Similarity: | 43.40% |
| Best Local Similarity: | 36.16% |
| Query Match: | 17.62% |
| DB: | 10 |
| US-09-763-712A-2_COPY_91_547 (1-457) x US-09-954-456-786 (1-5416) | |
| Qy 107 IleArgLeuAspSerValSerLeuArgMetClnGlnAspLeuMetArgSerArgLeuAsp | 126 |
| Db 2280 GTCCATTGGAGCGGAGGTCTTCTGCAGCCCCAGGCCCTGATGGAAACAAGGGTGAC | 2339 |
| Qy 127 ThrGluValAlaAsnLeuSerValIleMetClnGluMetLysLeuValAspSer- | 144 |
| Db 2340 CTGTTGTGGTGGTGGTCTGGGCACCTGCTGCATCTGGTTCCTAGTGACTCCCAGGAG | 2399 |
| Qy 145 -----LysHisGlyGlnLeuIleLysAsn | 152 |
| Db 2400 AGAGGGTGTCTGGCATACCTGGAGCAAGGAGGAAGAGGTGAACCTGGTCTCAGAG | 2459 |
| Qy 153 -----PheThrIleLeuGlnGlyPro | 159 |
| Db 2460 GTGAATTGGTAACCCTGGCAGAGATGGTCTCGTGGTGTCTCATGGTGTGTATA-GGGGCC | 2518 |
| Qy 160 ProGlyProArgGlyProArgGlyAspArg-----GlySerGlnGlyProProGly | 176 |
| Db 2519 CCTGGTCTCTGSAGCACAGGTGACCGGGGGCAAGCTGGGGCTGCTGGTCTGCTGGTGT | 2578 |
| Qy 177 ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAla | 196 |
| Db 2579 CCGTCTGGTCTCGGGGAAGCCCTGGTGTGAACGTGGCAGGTGCTGCTGCTGCCCAAC | 2638 |
| Qy 197 GlyGluArgGlyProIleGlyProAlaGlyProPro-----GlyGluArgGlyGly | 213 |
| Db 2639 GGATTTCTGGTCTGGGCTGTGTCTGGTGTCAACCGGGGTCTTAAGGAGAGAAGAGGCGC | 2698 |
| Qy 214 LysGlySerLysGlySerGln----- | 220 |

| | | | |
|--|----------|--|------|
| Db | 2699 | AAAGGGCCTAAGGCGTAAAAACGGTGTTGTGTGCCACAGAGCCCGCTTGGAGCTGCTGC | 2758 |
| Qy | 221 | -----GlyProLysGlySerArgGlySer- | 228 |
| Db | 2759 | CCAGCTGGTCCAATGGTCCCCCGGTCTGTGGGAAGTCGGTAGTGGAGGAGCCCGCT | 2818 |
| Qy | 229 | -----ProGlyLysProGlyProGlnGlyPro | 237 |
| Db | 2819 | GGTATGACTGGTTCCCTGGTCTGGACGGAGCTGGTCCCCCAGGACCCTCTGGTATT | 2878 |
| Qy | 238 | SerGlyAspProGlyProProGlyProGlyLysGluGlyLeuProGlyProGln--- | 256 |
| Db | 2879 | TCTGGCCCTCTCTGGTCCCCTGGTCTGTCTGGGAAAAGAGGCTTCGTGGTCTCGTGGT | 2938 |
| Qy | 257 | -----GlyProProGlyPheGln | 262 |
| Db | 2939 | GACCAAGGTCAGTTGGCCGAAGTAGGTGGAGTTGGTTCGCCCTCGCTTCGCT | 2998 |
| Qy | 263 | GlyLeuGln-----GlyThrValGlyGluProGlyValproGlyPro | 276 |
| Db | 2999 | GGTGAGAAGGGTCCCTCTGGAGAGGCTGGTACTGTGGACCTCTCGCACCTCCAGGTCCT | 3058 |
| Qy | 277 | ArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyPro ::: | 296 |
| Db | 3059 | CAGGGCTCTTCTGGTCTCTGGTATTCTGGGTCTCCCTGGCTCGAGAGGTGAACGTGGT | 3118 |
| Qy | 297 | ProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGluPro :: | 313 |
| Db | 3119 | CTACCTGGTGTTCGTGGTCTGGTGGTGAACCTGGTCTCTCTTGGCATTTGCCGGCCCTCT | 3178 |
| Qy | 314 | -----ThrProAlaProGluAspAsnGlyCysPro ::: | 323 |
| Db | 3179 | GGGCGCCGCTGGTCTCTGGTGTGGTGTAGTCTCTGGAGTCACGGGTGCTCTCT | 3232 |
| RESULT 10 | | | |
| US-09-880-107-2094 | | | |
| ; Sequence 2094, Application US/09880107 | | | |
| ; Patent No. US20020142981A1 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Horne, Darci T. | | | |
| ; APPLICANT: Vockley, Joseph G. | | | |
| ; APPLICANT: Scherf, Uwe | | | |
| ; APPLICANT: Gene Logic, Inc. | | | |
| ; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer | | | |
| ; FILE REFERENCE: 44921-5028-WO | | | |
| ; CURRENT APPLICATION NUMBER: US/09/880,107 | | | |
| ; CURRENT FILING DATE: 2001-06-14 | | | |
| ; PRIOR APPLICATION NUMBER: US 60/211,379 | | | |
| ; PRIOR FILING DATE: 2000-06-14 | | | |
| ; PRIOR APPLICATION NUMBER: US 60/237,054 | | | |
| ; PRIOR FILING DATE: 2000-10-02 | | | |
| ; NUMBER OF SEQ ID NOS: 3950 | | | |
| ; SOFTWARE: PatentIn Ver. 2.1 | | | |
| ; SEQ ID NO 2094 | | | |
| ; LENGTH: 5416 | | | |
| ; TYPE: DNA | | | |
| ; ORGANISM: Homo sapiens | | | |
| ; FEATURE: | | | |
| ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 J03464 | | | |
| US-09-880-107-2094 | | | |
| Alignment Scores: | | | |
| Pred. No.: | 2,67e-16 | Length: | 5416 |
| Score: | 444.00 | Matches: | 115 |
| Percent Similarity: | 43.40% | Conservative: | 23 |
| Best Local Similarity: | 36.16% | Mismatches: | 79 |
| Query Match: | 17.62% | Indels: | 102 |
| DB: | 10 | Gaps: | 10 |
| US-09-763-712A-2_COPY_91_547 (1-457) x US-09-880-107-2094 (1-5416) | | | |
| Ov | 107 | IleArgLeuAspSerValSerLeuArqMetGlnGlnAspLeuMetArgSerArgLeuAsp | 126 |

Qy 107 IleArqLeuAspSerValSerLeuArqMetGlnGlnAspLeuMetArqSerArqLeuAsp 126

```
Db 2280 GTCCATTATGGAGCGAGGTCCCTCTGGACCCCGAGGCGCTGATGAAACAGGCGTGAAC 2339
Qy 127 ThrGluValAlaAsnLysSerValIleMetGluGluMetLysLeuValAspSer----- 144
Db 2340 CTGGTGTGGTGGTCTGTGGCACTGCTGTGCTCATCTGGTCCATCTGGTCCAGTCCAGGAG 2399
Qy 145 -----LysHisGlyGlnLeuIleLysAsn 152
Db 2400 AGAGGGGTGCTGCTGGCATACCTGGAGGCAAGGAGAGAAAGGCTGAACCTGGTCTCAGAG 2459
Qy 153 -----PheThrIleLeuGlnGlyPro 159
Db 2460 GTGAATTTGGTAACCCCTGGCAGAGATGGTCTGCTGGTCTCATGGTCTGCTA-GGTGCC 2518
Qy 160 ProGlyProArgGlyProArgGlyAspArg-----GlySerGlnGlyProProGly 176
Db 2519 CCTGTCTCTGTGGAGCCACAGGTACCGGGCGAGCTGGGGCTGTGGTCTCTCTGCTGT 2578
Qy 177 ProThrGlyAsnLysGlyGlnLysGlyGluProGlyProGlyProGlyProAla 196
Db 2579 CCTGTGTGCTCCCGGGAAGCCCTGTGAACGTGCGAGGTGCTGTGGTCTGTGGCCCCAAC 2638
Qy 197 GlyGluArgGlyProIleGlyProAlaGlyProPro-----GlyGluArgGlyGly 213
Db 2639 GGATTGTGTGCTCCGGCTGTGCTGTCTGTCAACCGGCTGTCTAAAGAGAGAAAGAGAGGCG 2698
Qy 214 LysGlySerLysGlySerGln----- 220
Db 2699 AAAGGCGCTAAGGTTGAACACGGTGTGTGTGTCACACAGGCCCCGTTGGAGCTGCTGGC 2758
Qy 221 -----GlyProLysGlySerArgGlySer----- 228
Db 2759 CCAGCTGTGTCCAAATGTTCCCGCGTCTGTCTGGAAGTGTGGTGTGATGGAGGCCCCCT 2818
Qy 229 -----ProGlyLysProGlyProGlnGlyPro 237
Db 2819 GGTATGACTGTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2878
Qy 238 SerGlyAspProGlyProProGlyProGlyLysGluGlyLysGluGlyProGln--- 256
Db 2879 TCTGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2938
Qy 257 -----GlyProProGlyPheGln 262
Db 2939 GACCAAGTCCAGTTGGCGCACTGGAGAGTAGTGAGTTGGTGGTCCCGCTGCTGCTGCT 2998
Qy 263 GlyLeuGln-----GlyThrValGlyGluProGlyValProGlyPro 276
Db 2999 GGTGAGAGGGTCCCTCTGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3058
Qy 277 ArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProProGly 296
Db 3059 CAGGGTCTCTGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3118
Qy 297 ProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGluPro 313
Db 3119 CTACCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3178
Qy 314 -----ThrProAlaProGluAspAsnGlyCysPro 323
Db 3179 GGGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3232
```

RESULT 11

US-10-044-090-104

; Sequence 104, Application US/10044090

; Patent No. US20020137081A1

; GENERAL INFORMATION:

; APPLICANT: Olga Bandman

; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION

; FILE REFERENCE: PA-0028 US

; CURRENT APPLICATION NUMBER: US/10/044,090

; CURRENT FILING DATE: 2002-01-09

```
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 104
; LENGTH: 4149
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.19
; NAME/KEY: unsure
; LOCATION: 1492-1495, 1663, 1669, 4012, 4055
; OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-104
```

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 3,27e-16 | Length: | 4149 |
| Score: | 440.50 | Matches: | 95 |
| Percent Similarity: | 49.32% | Conservative: | 13 |
| Best Local Similarity: | 43.38% | Mismatches: | 58 |
| Query Match: | 17.48% | Indels: | 53 |
| DB: | 12 | Gaps: | 6 |

US-09-763-712A-2_COPY_91_547 (1-457) x US-10-044-090-104 (1-4149)

```
Qy 158 GlyProProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyPro 177
Db 1307 GGTCTGTGTGCTCGGGGAAGCCCTGTGTGAACGTGTGTGAGTGTGCTGCTGCTGCTGCT 1366
Qy 178 Thr-----Gly 179
Db 1367 AATGATTTGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1426
Qy 180 AsnLysGlyGlnLysGlyGluProGlyProGlyProGlyProGlyProAlaGlyGluArg 199
Db 1427 GCCAAGGCGCTAAGGTGAAACGGTGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1486
Qy 200 GlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyGlySerLysGlySer 219
Db 1487 GGGCCNNNGTCCAAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1546
Qy 220 GlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyPro----- 237
Db 1547 CTGTGTATGACTGGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1606
Qy 238 ---SerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGln 256
Db 1607 ATTTCTGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1666
Qy 257 -----GlyProProGlyPhe 261
Db 1667 GGNAGCAAGGACAGCAGCGCCGACCTGGAGAGTAGGAGCACCGGGTCCCCCTGCTGCTTC 1726
Qy 262 GlnGlyLeuGln-----GlyThrValGlyGluProGlyValProGly 275
Db 1727 GCTGTGAGAAAGGCTCCCTCTCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1786
Qy 276 ProArgGlyLeuProGlyValProGlyProGlyValProGlyMetProGlyProLysGlyPro 295
Db 1787 CTTAGGGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1846
Qy 296 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 312
Db 1847 GGTCTACCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1906
Qy 313 Pro-----ThrProAlaProGluAspAsnGlyCysPro 323
Db 1907 CTGGGGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1963
```

RESULT 12

US-09-880-107-3947

; Sequence 3947, Application US/09880107

; Patent No. US20020142981A1

; GENERAL INFORMATION:


```
; LOCATION: (5143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-206

Alignment Scores:
Pred. No.:      3.97e-16      Length:      5145
Score:          440.50       Matches:     93
Percent Similarity: 49.32%    Conservative: 13
Best Local Similarity: 43.38% Mismatches:   58
Query Match:     17.48%      Indels:     53
DB:              10         Gaps:        6

US-09-763-712A-2_COPY_91_547 (1-457) x US-09-925-299-206 (1-5145)

Qy 158 GlyProProGlyProArgGlyLeuProGlyProProGlyProAlaGlyGluArg 177
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2248 GTCTCCTGGTCTCTCGGGGAAGCCCTGTGTGAACGTGTGAGTGCTGCTGGCCCC 2307
Qy 178 Thr-----Gly 179
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2308 AATGGATTGCTGCTGCTGCTGCTGCTCAACCTGTGCTAAAGAGAAAAGAGA 2367
Qy 180 AsnLysGlyClnLysGlyGluProGlyProProGlyProAlaGlyGluArg 199
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2368 GCCAAAGGGCCTAATGGTGTGTTGTCGCCACAGGCCCGCTTGGAGCTGCT 2427
Qy 200 GlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySer 219
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2428 GGCCCAAGCTGCTCCAATGGTCCCCCGGCTCTGCTGGAAGTCGTGGTGATGGAGGCCCC 2487
Qy 220 GlnclyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyPro----- 237
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2488 CCTGGTAGTACTGGTTTCCTGCTGCTGCTGCACGAGCTGGTCCCACCAGCACCTCTGGT 2547
Qy 238 ----SerGlyAspProGlyProProGlyProProGlyLysGluGlyLeuProGlyProGln 256
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2548 ATTCTTGGCCCTCTCTGGTCCCCCTGGTCTGCTGGGAAAGAGGGCTTTCGTGGTCTCGT 2607
Qy 257 -----GlyProProGlyPhe 261
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2608 GTGACCAAGGTCCAGTTGGCGAACTGGAGAAAGTAGGTGTCAGTTGGTCCCCCTGGCTTC 2667
Qy 262 GlnclyLeuGln-----GlyThrValGlyLeuProGlyValProGly 275
||||| :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2668 GCTGGTGAGAAGGTCCTCTCGAGAGGCTGCTACTGTGGACCTCCTGGCACCTCCAGGT 2727
Qy 276 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 295
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2728 CCTCAGGGTCTTCTGGTGCTCTGTGTAATCTGGGTCTCCCTGGGTCGAGAGGCAACGT 2787
Qy 296 GlyProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGlu 312
||||| ||||| :||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2788 GGCTACAGAGTGTGCTGGTGTGGGTGAACCTGGTCTTCTGGCATTTCCCGGCCCT 2847
Qy 313 Pro-----ThrProAlaProGluAspAsnGlyCysPro 323
||||| :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2848 CCTGGGGCCGCTGGTCTCTCTGCTGTGTGGGTAGTCTCTGGAGTCAACGGTCTCCT 2904

RESULT 15
US-10-044-090-22
; Sequence 22, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE AC
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 5432
; TYPE: DNA
```

ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc_feature
 OTHER INFORMATION: Incyte ID No. US20020137081A1 1383093.13
 US-10-044-090-22

Alignment Scores:
 Pred. No.: 4,16e-16 Length: 5432
 Score: 440.50 Matches: 114
 Percent Similarity: 43.08% Conservative: 23
 Best Local Similarity: 35.85% Mismatches: 80
 Query Match: 17.48% Indels: 102
 DB: 12 Gaps: 10

US-09-763-712A-2_COPY_91_547 (1-457) x US-10-044-090-22 (1-5432)

QY 107 IleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeuMetArgSerArgLeuAsp 126
 Db 2283 GTCCATTGGAGCGGAGGTCCTCTTGGACCCCGGCTGATGGAAACAAGGGTGAAC 2342
 QY 127 ThrGluValAlaAsnLeuSerValIleMetGluGluMetLysLeuValAspSer----- 144
 Db 2343 CTGGTGTGGTGTGTGTGGGCACTGCTGCTCCATCTGCTAGTGGACTCCACAGGAG 2402
 QY 145 -----LysHisGlyGlnLeuLeuLysAsn 152
 Db 2403 AGAGGGGTGCTGCTGCATACCTGGAGGCAAGGAGGAGGAGGAGGAGGAGGAGGAG 2462
 QY 153 -----PheThrIleLeuGlnGlyPro 159
 Db 2463 GTGAATTGGTAACCTGGCAGAGATGGTGTGGTGTCTCTGGTGTCTGTA-GGTGCC 2521
 QY 160 ProGlyProArgGlyProArgGlyAspArg-----GlySerGlnGlyProGly 176
 Db 2522 CCTGGTCTGTGGAGCCACAGGTACCGGGGAGGCTGGGGCTGCTGCTGCTGCTGCT 2581
 QY 177 ProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProAla 196
 Db 2582 CCTGTGTGGTCTCGGGGAAAGCCCTGGTGAACGTGGTGGTGGTGGTGGTGGTGGT 2641
 QY 197 GlyGluArgGlyProIleGlyProAlaGlyProPro-----GlyGluArgGlyGly 213
 Db 2642 GGATTTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2701
 QY 214 LysGlySerLysGlySerGln-----220
 Db 2702 AAAGGCTTAAGGGTGAAGAGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2761
 QY 221 -----GlyProLysGlySerArgGlySer-----228
 Db 2762 CCAGTGTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2821
 QY 229 -----ProGlyLysProGlyProGlnGlyPro 237
 Db 2822 GGTATGACTGGTTTCCCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2881
 QY 238 SerGlyAspProGlyProProGlyProGlyLysGluGlyLeuProGlyProGln--- 256
 Db 2882 TCTGGCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2941
 QY 257 -----GlyProProGlyPheGln 262
 Db 2942 GACCAAGTTCAGTTGGCGAAGTGGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3001
 QY 263 GlyLeuGln-----GlyThrValGlyGluProGlyValProGlyPro 276
 Db 3002 GTTGAGAGGGTCCCTCTGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3061
 QY 277 ArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGly 296
 Db 3062 CAGGGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3121
 QY 297 ProProGlyProSerGlyAlaVal-----ValProLeuAlaLeuGlnAsnGluPro 313

Db 3122 CTACCAGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3181
 QY 314 -----ThrProAlaProGluAspAsnGlyCysPro 323
 Db 3182 GGGGGCCGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3235

Search completed: March 21, 2003, 12:32:08
 Job time : 248.049 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 08:08:41 ; Search time 1510.54 Seconds
(without alignments)
4899.781 Million cell updates/sec

Title: US-09-763-712A-2_COPY_91_547
Perfect score: 2520
Sequence: 1 MNSOLNSTGQMENTTISQ.....EDVNFNICEKDKRETVLSSAL 457

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n_model -DEV=xlp
-O=/cgn2_1/USPRO_pool/US09763712/runat_14032003_100950_18109/app_query.fasta_1.1877
-DB=EST -QFMT=tastap -SUFF1=first -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712@cgn_1.1.4575 -runat_14032003_100950_18109 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-FGAPOP=10 -FGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|----------|-------------|
| 1 | 2344 | 93.0 | 3305 | 11 | BC009162 | Mus muscu |
| 2 | 1460.5 | 58.0 | 957 | 14 | BQ955927 | AGENCOURT |
| 3 | 1256 | 49.8 | 861 | 14 | BQ713873 | AGENCOURT |
| 4 | 1251 | 49.6 | 906 | 13 | B1456109 | 603172765 |
| 5 | 1239.5 | 49.2 | 936 | 14 | BQ891432 | AGENCOURT |
| 6 | 1022 | 40.6 | 580 | 12 | BE910803 | 601661855 |
| 7 | 1007 | 40.0 | 668 | 14 | BQ674807 | AGENCOURT |
| 8 | 957 | 38.0 | 601 | 10 | BE290299 | 601089246 |
| 9 | 916.5 | 36.4 | 967 | 9 | AL568743 | AL568743 |
| 10 | 842 | 33.4 | 552 | 10 | AW958053 | EST370123 |
| 11 | 831 | 33.0 | 638 | 10 | BB248064 | BB248064 |
| 12 | 798.5 | 31.7 | 884 | 14 | BQ934501 | AGENCOURT |
| 13 | 767 | 30.4 | 808 | 14 | BQ771366 | BQ771366 |
| 14 | 669 | 26.5 | 500 | 14 | BM676508 | UI-E-EJ0- |
| 15 | 663 | 26.3 | 507 | 14 | BM713891 | UI-E-EJ0- |
| 16 | 627.5 | 24.9 | 591 | 13 | B1445884 | da133d10. |
| 17 | 601.5 | 23.9 | 1082 | 14 | BM907108 | AGENCOURT |
| 18 | 600 | 23.8 | 357 | 10 | AW379436 | CM4-HT024 |
| 19 | 594 | 23.6 | 350 | 9 | AA012704 | RPUI402CG |
| 20 | 588 | 23.3 | 339 | 9 | AA304251 | EST17006 |
| 21 | 560 | 22.2 | 517 | 9 | AL543000 | AL543000 |
| 22 | 538 | 21.3 | 581 | 14 | BQ127513 | BQ127513 |
| 23 | 537 | 21.3 | 310 | 9 | AA361740 | EST71069 |
| 24 | 516 | 20.5 | 465 | 10 | AW240221 | up30b07.Y |
| 25 | 503.5 | 20.0 | 839 | 13 | BI697412 | 603348154 |
| 26 | 478 | 19.0 | 943 | 14 | BQ887163 | AGENCOURT |
| 27 | 477 | 18.9 | 679 | 14 | W27610 | W27610 |
| 28 | 460 | 18.3 | 441 | 10 | BB749743 | BB749743 |
| 29 | 458.5 | 18.2 | 569 | 10 | AV616076 | AV616076 |
| 30 | 457.5 | 18.2 | 1029 | 17 | CNS04VYG | AG1309697 |
| 31 | 456.5 | 18.1 | 792 | 13 | BI456015 | 603170460 |
| 32 | 456 | 18.1 | 420 | 12 | BE929813 | RC3-GN004 |
| 33 | 455.5 | 18.1 | 896 | 14 | BQ955616 | AGENCOURT |
| 34 | 453.5 | 18.0 | 877 | 12 | BG681943 | BG681943 |
| 35 | 452 | 17.9 | 688 | 13 | BI249235 | 602995511 |
| 36 | 452 | 17.9 | 3951 | 11 | BC013626 | Mus muscu |
| 37 | 451.5 | 17.9 | 921 | 14 | BQ921102 | AGENCOURT |
| 38 | 451.5 | 17.9 | 1192 | 14 | BM905999 | AGENCOURT |
| 39 | 449.5 | 17.8 | 907 | 14 | BQ222937 | AGENCOURT |
| 40 | 448 | 17.8 | 946 | 14 | BQ942234 | AGENCOURT |
| 41 | 447 | 17.7 | 862 | 12 | BG171919 | 602333020 |
| 42 | 446.5 | 17.7 | 608 | 13 | BM425818 | pgf2c.pk0 |
| 43 | 446 | 17.7 | 651 | 10 | BB613976 | BB613976 |
| 44 | 446 | 17.7 | 819 | 12 | BG679800 | 602626481 |
| 45 | 446 | 17.7 | 1120 | 14 | BM809933 | AGENCOURT |

ALIGNMENTS

RESULT 1
BC009162
LOCUS
DEFINITION BC009162 3305 bp mRNA linear HFC 07-AUG-2002
Mus musculus, similar to scavenger receptor with C-type lectin,
clone IMAGE:2811487, mRNA.
ACCESSION BC009162
VERSION BC009162.1 GI:14714370
KEYWORDS HTC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 3305)
AUTHORS Strausberg, R.
TITLE Direct Submission

JOURNAL

Submitted (05-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs@remai.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK plate: 5 Row: h Column: 10
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis

This clone has the following problem: incomplete processing.

FEATURES
source

Location/Qualifiers
1. 3305
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="IMAGE:2811487"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."
/clone_lib="NCL CGAP_Mam5"
/clone_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT 1031 a 757 c 760 g 757 t
ORIGIN

Alignment Scores:

Pred. No.: 6,95e-150 Length: 3305
Score: 2344.00 Matches: 417
Percent Similarity: 96.50% Conservative: 24
Best Local Similarity: 91.25% Mismatches: 16
Query Match: 93.02% Indels: 0
DB: 11 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x BC009162 (1-3305)

Qy 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
Db 951 ATGAATAGCCAGCTCAGCTCATTCACAGGTGAGTGGACAACTATACCACTATCTCACAG 1010
Qy 21 AlaAsnGluAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
Db 1011 GCCAACGAGCAGAGCTGAAAGACCTTCAGGACTTACACAAAGATACAGAAAATAGAACA 1070
Qy 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
Db 1071 GCTGTCAAGTTCAGCAACTTCAGGAACGCTTCCAGGCTTTTCAGACAGATATGTGAAC 1130
Qy 61 IleIleSerAsnIleSerTyrThrAlaHisLysLeuArgThrLeuThrSerAsnLeuAsn 80
Db 1131 ATCATTTAGCAACATCAGCTACAGACCCCATCATCTGCGGACACTGACCACTCTGAAT 1190
Qy 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
Db 1191 GATGTTAGGACCATGTCAGACACACCTTGACACAGACACACGATGACCTGACCTCCTTG 1250
Qy 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnAspLeu 120
Db 1251 AATAACACACTAGTCAACATCGCTTGGATTCTATTTCTCTCAGGATGACGACGACATG 1310

Qy 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetLys 140
Db 1311 ATGAGGTCAAACTTAGACACTGAAGTGGCCAACTTATCATAGTGGTATGGAAGAGATGAAA 1370
Qy 141 LeuValAspSerLysHisGlnLeuLysAsnPheThrIleLeuGlnGlyProPro 160
Db 1371 CTGGGTGACTCCAAAGCAGCGTCAGCTCATCAAGAACTTTACCATTTCTACAAGGTCTCTCT 1430
Qy 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGlyAsn 180
Db 1431 GCCCCAGAGGTCCAAAGGTGACAGAGGATCTCAGGAGACCACTGGTCCCACTGGCAAC 1490
Qy 181 LysGlyGlnLysGlyGlnLysGlyGluProGlyProGlyProGlyProAlaGlyGluArgGly 200
Db 1491 AAGGACAGAAAGGAGAGAGGAGAGCGCTGCTCCACCTGGCCCTCGCGGTGAGAGGGGC 1550
Qy 201 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
Db 1551 ACAATTGGACCACTCGGCCCTCTCGAGAGCGTGGCAGCAAAAGGATCCAAAGGCTCACAG 1610
Qy 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
Db 1611 GGTCCAAAGGATCTGCTGGGTCCCGAGGGAAGCCCTGAGGACCTAGTGGGGAC 1670
Qy 241 ProGlyProProGlyProGlyLysGlyGluLeuProGlyProGlyProGlnGlyProProGly 260
Db 1671 CCAGGACCACTGAGTCCAGGCAAGGATGAGCTCCCTGGCCCTCAGGCGCTCTCTGGC 1730
Qy 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
Db 1731 TTCCAGGGACTACAGGGCACTGCGGTGAGCCTGGAGTACCTGGACCTCGGGGTGGCA 1790
Qy 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProProGlyPro 300
Db 1791 GCTTGGCCAGGGGTGCGAGCATGCTGGGCTTAAGGAGCACTTGGCCCTCAGGCCCC 1850
Qy 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
Db 1851 TCAGGAGCAATGGAGCCATTGCTCTGCAGAAATGAACCAACCCAGCATCAGAGGTCAAC 1910
Qy 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTrpPheSerValGluLys 340
Db 1911 GGATGTCCACCTCACTGGAAGAACTTCACAGATAAATGCTACTATTTTCATTGGAAAAA 1970
Qy 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
Db 1971 GAAATTTTGAAGATGCTAAGCTTTTCTGTGAAGACAAATCTTCCCATCTCGTTTCATA 2030
Qy 361 AsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
Db 2031 AACTCAAGAGAAAGACAGCAATGGATAAAAGCATACCGTGGGAGAGAAAGCCATTGG 2090
Qy 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
Db 2091 ATCGGCTCACAGACTCAGAACAGAAAGCAATGGAAGTGGCTAGACGGGTCACTGTT 2150
Qy 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
Db 2151 GATTACAAAACCTGGAAGCTGGCAACCAAGATACCTGGGCGAGTGGCCATGGGCAGGA 2210
Qy 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
Db 2211 GAAGACTGTGCTGGCTGATTACGACGACAGTGAATGACTTCCAGTGTGATGAATC 2270
Qy 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
Db 2271 AATAACTTCATTGTGAGAAGAAAGGAGGAGGAGTACCATCATCATATTA 2321
RESULT 2
BQ955927
LOCUS
DEFINITION AGENCOURT_8763247 NIH_MGC_129 Mus musculus cdna clone IMAGE:6313500
957 bp mRNA linear EST 21-AUG-2002

5', mRNA sequence.
BQ955927
VERSION BQ955927.1 GI:22371405
EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 957)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@remail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13740 row: 1 column: 13
High quality sequence stop: 665.

FEATURES
source Location/Qualifiers
1..957
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6313500"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Olfactory epithellum; Vector:
pcmw-SPORT6.1.ccdB; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dt. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH_MGC Library."
BASE COUNT 260 a 269 c 258 g 169 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 3,26e-90 Length: 957
Score: 1460.50 Matches: 284
Percent Similarity: 93.73% Conservative: 15
Best local Similarity: 89.03% Mismatches: 13
Query Match: 57.96% Indels: 8
DB: 14 Gaps: 3

US-09-763-712A-2_COPY_91_547 (1-457) x BQ955927 (1-957)

| | | | | |
|----|-----|---|--|----|
| QY | 2 | AsnSerGln--- | LeuAsnSerPheThrGlyClnMetGluAsnIleThrThrIleSerGln | 20 |
| | | | | |
| Db | 8 | AATTCCGGGATCTTCAGCTCATTTACAGAGTCAGATGACAACATTACCACATATCTCACAG | 67 | |
| QY | 21 | AlaAsnClnGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr | 40 | |
| | | | | |
| Db | 68 | GCCAACGACGAGACCTCGAAGACCTTCAGACTTCACAGGATGATACAGAAATAGAACCA | 127 | |
| QY | 41 | AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValasn | 60 | |
| | | | | |
| Db | 128 | GCTGTCAAGTTACGCCAACTTGAGGAACGCTCCAGCTCTTGAGACAGATATGTGAAC | 187 | |
| QY | 61 | IleIleSerAsnIleSerTyThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn | 80 | |
| | | | | |
| Db | 188 | ATCATTAGCAACATCAGCTACAGCCCATCACCTGAGGACACTGACGACCAATCTGAAT | 247 | |
| QY | 81 | GluValArgThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu | 100 | |
| | | : : : | : : : | |
| Db | 248 | GATGTAGGACCACATGCACAGACACCTTCACACAGACACACGGATGACCTGCCTCTTG | 307 | |
| QY | 101 | AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu | 120 | |
| | | | | |
| Db | 308 | AATAACACACATAGTCAACATCCGCTGGATTCTATTCTCTCAGGATGACCAACAGCATG | 367 | |
| QY | 121 | MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys | 140 | |

/note="Organ: olfactory epithelium; Vector: pcMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library." 147 t 1 others

BASE COUNT 221 a 253 c 239 g 147 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 2,52e-76 Length: 861
Score: 1256.00 Matches: 229
Percent Similarity: 91.01% Conservativeness: 14
Best Local Similarity: 85.77% Mismatches: 20
Query Match: 49.84% Indels: 4
DB: 14 Gaps: 0

US-09-763-712A-2_COPY_91_547 (1-457) x BQ713873 (1-861)

QY 120 LeuMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMet 139
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 13 ATGATGAGGTCAAGGTAGACACTGAAGTGGCCAACTTATCAGTGGTATGGAAGAGATG 72
QY 140 LysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhetHrIleLeuGlnGlyPro 159
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 73 AAATGGTGGTCTCCAAAGCAGCGTCTCAGCTCATCAAGAACTTTTACCATTCTCAAGGTCT 132
QY 160 ProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGly 179
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 133 CTGGCCCCAGAGTCCAAAGAGGTGACAGAGGATCTCAGGACCACCTGGTCCAACTGGC 192
QY 180 AsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGlyGluArg 199
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 193 AACAGGGACAGAAAGGAGAGAGGAGCGCTGTGCTCCACCTGGCCCTCGGGGTGAGAGG 252
QY 200 GlyProIleGlyProAlaGlyProProGlyGluArgGlyGlyGlySerLysGlySer 219
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 253 GGCACAAATTTGGACCTGCTGGAGAGCGCTGGAGCAAGGATGCCAAAGAGCTCA 312
QY 220 GlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGly 239
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 313 CAGGGTCCCAAGGATCTCGTGGTCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 372
QY 240 AspGlyProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProPro 259
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 373 GACCCAGGACCCAGGTCCAGCAGGCAAGGATGGACTCCCTGGCCCTCAGGCGCTCCT 432
QY 260 GlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeu 279
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 433 GCGTCCAGGGACTACAGGGCACTGTGGGTGAGCCTGGAGTACCTGGACCTCGGGGGTTG 492
QY 280 ProGlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGly 299
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 493 CCAGGCTTCCAGGGGTGCCAGGATCCCTGGCCCTAAGGGACCACTGTGGCCCTCCAGGC 552
QY 300 ProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAsp 319
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 553 CCCTCAGGAGCAATGAGGCATTTGGCTCTGCAAAATGAACCAACCCCAACATCAGAGTGC 612
QY 320 AsnGlyCysProProHisTrpLysAsnPhetHrAspLysCysTrpThrPheSerVal-cl 339
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 613 ACAGGATGTCCGCTCCTGGAAGAACTTTCACAGATAAATGCTACTATTTTTCCTTGGCA 672
QY 339 uLysGluIlePheGluAspAlaLysLeu-PheCysGluAspLysSerHisLeuValP 359
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 673 AAAAAAATTTTCGAAAAATGCTACCTTTTCTGCGAACAATAATCTCCCCATCCGGGT 732
QY 359 heIleasnThrArgGluGluGlnInTrp-Ile-LysLysGlnMetValGlyArgGluSe 378
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 733 TCATACCCCCCAAGAAAAACAGCATGGGATAAAAAACCCCTACCTTCGGGGGACCAAGAAC 792
QY 378 rHisTrpIleGly 382
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 793 CCGCTGGATCGGG 805

RESULT 4
BI456109

LOCUS
DEFINITION

ACCESSION
VERSION

KEYWORDS
SOURCE

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

CONTACT: Robert Strausberg, Ph.D.
Email: cgsbbs-remail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
DNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

High quality sequence stop: 730.

Location/Qualifiers

1..906

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:5251888"

/clone_lib="NCL_CGAP_Mam5"

/tissue_type="tumor, gross tissue"

/dev_stage="7 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pcMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 247 a 228 c 256 g 175 t

ORIGIN

Alignment Scores:

Pred. No.: 5,8e-76 Length: 906

Score: 1251.00 Matches: 239

Percent Similarity: 89.12% Conservativeness: 15

Best Local Similarity: 83.86% Mismatches: 27

Query Match: 49.64% Indels: 6

DB: 13 Gaps: 2

US-09-763-712A-2_COPY_91_547 (1-457) x BI456109 (1-906)

QY 76 ThrSerAsnLeuAsnGluValArgThrThrCysThrAspThrLeuThrLysHisThrAsp 95
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 3 ACCGCAATCTCAATGATGTTAGGACCACATGCACAGACACCTTGACACACACGAT 62
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

QY 96 AspLeuThrSerLeuAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArg 115
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 63 GACCTGACCTCTTGAATAACACACTAGTCAACATCCGCTGGATTTCTTCTCTCAGG 122
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

QY 116 MetGlnGlnAspLeuMetArgSerArgLeuAspThrGluValAlaAsnSerValIle 135
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 123 ATGACGACGACATGATGAGTCAAGGTAGACACTGAAGTGGCCCAACTATCAGTGGT 182
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

QY 136 MetGluGluMetLysLeuValAspSerLysHisGlyGlnLeuIleLysAsnPhetHrIle 155
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 183 ATGGAAGAGATGAACCTGGTTCCTCCAGACGCGTCACTCATCAAGAACTTTACCAT 242
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

QY 156 LeuGlnGlyProGlyProArgGlyProArgGlyAspArgGlySerGlnGlyProPro 175
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

```

|||||
Db 243 CTACAGGTCTCTCTGGCCCCAGAGTCCAAAGGTGACAGAGGATCTCAGGAGCACCT 302
QY 176 GlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyProGlyPro 195
Db 303 GTCTCACTGGCAACAGGAGCAAGAGAGAGAGAGGAGGAGCCCTGGTCCACTGGCCCT 362
QY 196 AlaGlyGluArgGlyProIleGlyProAlaGlyProGlyProGlyGluArgGlyGlyLysGly 215
Db 363 GCGGTGAGAGGGGACAAATGGACCAAGTCAGTCGCCCTCTCTGGAGAGCGTGGCAGCAAGGA 422
QY 216 SerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGlyProGln 235
Db 423 TCCAAAGGCTACAGGGTCCCAAGAGATCTCGTGGGTCCCCAGTAGGAGCTGGCCCTCAA 482
QY 236 GlyProSerGlyAspProGlyProGlyProGlyProGlyLysGluGlyLeuProGlyPro 255
Db 483 GGACTAGTGGGGAGCCAGGACCAAGTCCAGGAGGATGGAGTCCCTGGCCCT 542
QY 256 GlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGly 275
Db 543 CAGGCGCTCTCTGGCTTCCAGGAGTACAGGCGACTGTGGGTGAGCTGGAGTACCTGGA 602
QY 276 ProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGlyProPro 295
Db 603 CCTCGGGGTTCACAGGCTTCCAGGGTGGCCAGGATGCCCTGGGCTTAAGGGAGCACT- 661
QY 296 GlyProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGluProThrPro 315
Db 662 AGCCCTCAGGCGCTCAGGAGCATG-GAGCCATGGTTCGAGAAATGAACCAAGCCCA 720
QY 316 AlaProGlu--AspAsnGlyCysProProHisTrpLysAsnPheThrAsp---LysCysT 334
Db 721 GCATCAGAGGTTCACGCGATGTCCTCTCACTGCGAAGAACTTTCACGAGTAAATGGCT 780
QY 334 yTyrPheSerVal---GluLysGluIlePheGluAspAlaLysLeuPheCysGluAspL 353
Db 781 ACTATTTTCCATTCACAAAAGAAATTTTGAAGATTGCTTAAGCGTTCTCTGGAGAA 840
QY 353 ySerSerHis 356
Db 841 AATTTTCCAAT 851

RESULT 5
BO991432
LOCUS
DEFINITION BO991432 936 bp mRNA linear EST 16-AUG-2002
5', mRNA sequence.
AGENCOURT_8754195 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6332754
BO991432
VERSION BO991432.1 GI:22283446
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 936)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13790 row: n column: 19
High quality sequence start: 25
High quality sequence stop: 632.
Location/Qualifiers

```

```

source
1..936
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6332754"
/clone_lib="NIH_MGC_130"
/lab_host="DH10B (phage-resistant)"
/note="Organ: otocysts; Vector: pCMV-SPORT6.1.cdb;
Site_1: EcorV; Site_2: NotI; Cloned unidirectionally.
Primer: Oligo dt. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a
NIH_MGC Library."
BASE COUNT 273 a 250 c 244 g 163 t 6 others
ORIGIN
Alignment Scores:
Pred. No.: 3,64e-75 Length: 936
Score: 1239.50 Matches: 241
Percent Similarity: 94.49% Conservative: 16
Best Local Similarity: 88.60% Mismatches: 9
Query Match: 49.19% Indels: 6
DB: 14 Gaps: 1
US-09-763-712a-2_COPY_91_547 (1-457) x BO991432 (1-936)
QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrLeSerGln 20
Db 131 ATGAATAGCCAGCTCAGCTCATTCACAGGTCAGATGGCAACATTACCACTATCTCACAG 190
QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
Db 191 GCCACGACGAGCGCTGAAGACCTTCAGGACTTACACAGGATACAGAAATAGAACCA 250
QY 41 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 60
Db 251 GCTGTCAAGTTCAGCCAACTTCAGGAACTTCAGGAGCTTCAGACAGATATTGTGAAC 310
QY 61 IleLeSerAsnIleSerTyrThrAlaHisLysLeuArgThrLeuThrSerAsnLeuAsn 80
Db 311 ATCATTAGCAACATCAGCTACACAGCCATCACCTGAGGACACTGACCAATCTGAAT 370
QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
Db 371 GATGTTAGGACCAATGACAGACACTTGCACAGACACAGGATGACCTGACCTCTCTG 430
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
Db 431 ATAACACACTAGTCAACATCCGCTGGATTTCTTCTCAGGATGAGCAAGACATG 490
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
Db 491 ATGAGGTCAAAGTTAGACACTGAAGTGGCCAACTTATCAGTGGTTATGAAGAGATGAAA 550
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
Db 551 CTGGTTGACTCAAGCAGCGCTCAGCTCAGTCAAGAACTTTACCAATCTCAAGGTCCTCT 610
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProGlyAsn 180
Db 611 GSCCCAGAGGTCCAAAGAGGTGACAGAGATCTCAGGAGCCACTGCTGCTCAACTGGCAAC 670
QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGlyGluArgGly 200
Db 671 AAGGGACAAAGGAGAGAGAGGAGAGCGCTGGTCCACCTGGCCCTGCGGGTGAGAGGGGC 730
QY 201 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyGlySerLysGlySerGln 220
Db 731 ACAATTTGGACCAGTCCGCCCTCTCTGGAGAGCGTGGCAGCAAGAAAGGATCCAAAGGCTCACAG 790
QY 221 Gly-ProLysGlySerArgGlySer--ProGlyLysProGlyProGlnGlyProSerGlyA 240
Db 791 GGNTCCAAAGGGATCTCGTGGGTGCCCGGAGGAGGCTGGCCCTCAAGGAGACTAGTGGGG 850
QY 240 spProGly-ProProGlyProProGly-LysGluGlyLeuPro-GlyProGlnGlyProP 259

```

```

||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 851 ACCAGNACCAGTCCACAGNCAGGATGAGTACCTCGCTGGCCCTCCTAGGNCNT 910
QY 259 ro---GlyPheGlnGlyLeuGlnGly 266
Db 911 CCTGGCTCNCAGGGGACTACAGGG 936

RESULT 6
BE910803 BE910803 580 bp mRNA linear EST 29-SEP-2000
LOCUS 601661855F1 NCI_CGAP_Maml Mus musculus cDNA clone IMAGE:3962292 5',
DEFINITION mRNA sequence.
ACCESSION BE910803
VERSION BE910803.1 GI:10407765
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LL9128 row: m column: 13
High quality sequence stop: 580.
Location/Qualifiers
1..580
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3962292"
/clone_lib="NCI CGAP Maml"
/tissue_type="tumor, biopsy sample"
/dev_stage="10 months, virgin"
/lab_host="DH10B"
/Note="Organ: Mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 159 a 153 c 161 g 107 t
ORIGIN

Alignment Scores:
Pred. No.: 1..5e-60 Length: 580
Score: 1022.00 Matches: 177
Percent Similarity: 96.34% Conservative: 7
Best Local Similarity: 92.67% Mismatches: 7
Query Match: 40.56% Indels: 0
DB: 12 Gaps: 0

US-09-763-712A-2_COPY_91_547 (1-457) x BE910803 (1-580)

QY 214 LysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysProGly 233
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 6 AAAGATCAAAAGGCTCACAGGTCGCCAAAGGATCTCGTGGTCCCGAGGAAGCCTGGC 65
QY 234 ProGlnGlyProSerGlyAspProGlyProGlyProGlyProGlyLysGlnGlyLeuPro 253
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 66 CCTCAGGACCTAGTGTGGGACCCAGGACCCAGGTCACAGGGAAGGATGGACTCCCT 125
QY 254 GlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGlyVal 273
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 126 GGCCCTCAGGGCCCTCCTGGCTTCAGGGGACTACAGGGGACTGTGGGTGAGCCTGGAGTA 185

```

```

QY 274 ProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLysGly 293
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 186 CTGGACCTCGGGGTTGCCAGGCTTCCAGGGGTCCAGGATGCTGGCGCTAAGGA 245
QY 294 ProProGlyProProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluPro 313
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 246 CCACCTGGCCCTCCAGGCCCTCAGGAGCAATGGAGCCCATTTGGCTCTGCAGATGAACA 305
QY 314 ThrProAlaProGluAspAsnGlyCysProProHisTrpLysAsnPheThrAspLysCys 333
||||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 306 ACCCAGCATCAGAGTCAACGGATGTCGCCCTCACTGGAAGAACTTCACAGATAATGC 365
QY 334 TyrTyrPheSerValGluLysGluLeuPheGluAspAlaLysLeuPheCysGluAspLys 353
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 366 TACTATTTTTCATTTGAAAAAGAAATTTTGAAGATGCTAAGCTTTTCTGTGAAGACAAA 425
QY 354 SerSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysGlnMet 373
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 426 TCTTCCCATCTCGTTTTCATAAATCAAGAGAGAAGACGCAATGGATAAAAAAGCATACC 485
QY 374 ValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLys 393
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 486 GTGGGAGAGAAAGCCATTTGGATCGGCCCTCACAGACTCAGACAGCAAGCAATGGAAG 545
QY 394 TrpLeuAspGlyThrSerProAspTyrLysAsn 404
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 546 TGGCTAGAGGGGTACCTGTTGATTACAAAAA 578

RESULT 7
BQ674807 BQ674807 668 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8035172 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6212672
DEFINITION 5', mRNA sequence.
ACCESSION BQ674807
VERSION BQ674807.1 GI:21785641
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 668)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LL92377 row: c column: 09
High quality sequence stop: 656.
Location/Qualifiers
1..668
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6212672"
/clone_lib="NIH_MGC_102"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: salivary gland; Vector: pOT87; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 216 a 148 c 167 g 137 t
ORIGIN

```


RESULT 9

AL568743/c
 LOCUS AL568743 LTI_FL002_PL1 Homo sapiens cDNA clone CS0DE005YH04 3 prime
 DEFINITION mRNA sequence.
 EST 16-FEB-2001
 ACCESSION AL568743
 VERSION AL568743.1 GI:129233387
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 967)
 AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source
 1. .967
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DE005YH04"
 /clone_lib="LTI_FL002_PL1"
 /lab_host="DH10B"
 /note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand
 cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-stranded cDNA was digested with Not I
 and cloned into the Not I and Eco RV sites of the
 pCMVSPORT 6 vector. Library was constructed by life
 Technologies. Contact : Feng Liang Life Technologies, a
 division of Invitrogen 9800 Medical Center Drive Rockville
 , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
 fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 222 a 221 c 217 g 290 t 17 others
 ORIGIN

Alignment Scores:

Pred. No.: 3 82e-53 Length: 967
 Score: 916.50 Matches: 174
 Percent Similarity: 91.1% Conservative: 1
 Best Local Similarity: 90.62% Mismatches: 14
 Query Match: 36.37% Indels: 5
 DB: 9 Gaps: 3

US-09-763-712a-2_COPY_91_547 (1-457) x AL568743 (1-967)

QY 266 GlyThrValGlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyVal 285
 |||||
 Db 947 GGCACCGTTGGGAGCGTGGTCT---GGACCTCGGGACTGCCAG---CTTGCCTGGGTA 894
 QY 286 ProGlyMetProGlyProGlyProGlyProGlyProGlyProSerGlyAlaValVal 305
 |||||
 Db 893 CCAGGATCCAGCCCAAG---CCCGCCCGCCCTCTCTCC-CCATCAGGAGCGGTGTG 838
 QY 306 ProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProProHis 325
 |||||
 Db 837 CCCTGSCCTCGAGATGAGCAACCCCGGACCCGGAGACAATGGCTGCCGCGCTCAC 778
 QY 326 TrpLysAsnPhetrAspLysCysTyrTyrrPheSerValcLulysGluIlePheGluAsp 345
 |||||
 Db 777 TGSAGAACTTCACACAAATGCTACTATTTTCAGTTGAGAAAGAAATTTTGGAGAT 718
 QY 346 AlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGlu 365
 |||||
 Db 717 GCAAGCTTTCTGTGAAGACAAGTCTCACATCTTGTGTTTTCATAAACACTAGAGAGAA 658
 QY 366 GlnGlnTrpIleLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAsp 385
 |||||
 Db 657 CAGCAATGATATAAAACAGATGGTAGGAGAGAGCCACTGGATCGGCCTCACAGAC 598

QY 386 SerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrIlysAsnTrp 405
 |||||
 Db 597 TCAGACCGTGAAATCAATGGAAGTGGCTGGATGGACATCTCCAGACTACAAAAATGG 538
 QY 406 LysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGly 425
 |||||
 Db 537 AAAGCTGGACAGCCGATAACTGGGCTCATGC-CATGGGCCAGGAGAAGACTGTGCTGG 479
 QY 426 LeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCys 445
 |||||
 Db 478 TTGATTATTGCTGGGAGTGGAGACGATTTCCAAATGTGAAGACGTCATTAATTCTTCATTC 419
 QY 446 GluLysAspArgGluThrValLeuSerSerAlaLeu 457
 |||||
 Db 418 GAAAAAGACAGGAGAGAGACTGTCTATCTGCATTA 383
 RESULT 10
 AW958053 552 bp mRNA linear EST 01-JUN-2000
 LOCUS EST370123 MAGe resequences, MAGe Homo sapiens cDNA, mRNA sequence.
 DEFINITION AW958053
 ACCESSION AW958053
 VERSION AW958053.1 GI:8147736
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 552)
 AUTHORS Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt
 , I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
 Quackenbush, J.
 TITLE Assessment of gene expression patterns in a model of colon tumor
 metastasis using a 19,200 element cDNA microarray
 JOURNAL Unpublished (2000)
 COMMENT Contact: John Quackenbush
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 3528
 Fax: 301 838 0208
 Email: johnq@tigr.org
 Plate: 115
 Seq primer: Reverse.
 Location/Qualifiers
 1. .552
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MAGe resequences, MAGe"
 /note="Vector: pBluescriptSKm"

BASE COUNT 176 a 112 c 143 g 121 t
 ORIGIN

Alignment Scores:

Pred. No.: 2 63e-48 Length: 552
 Score: 842.00 Matches: 149
 Percent Similarity: 98.68% Conservative: 1
 Best Local Similarity: 98.03% Mismatches: 1
 Query Match: 33.41% Indels: 1
 DB: 10 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x AW958053 (1-552)

QY 307 LeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsnGlyCysProHisTrp 326
 |||||
 Db 1 CTGGCCCTCAGAAATCAGCCACCCCGGACCGGAGACAATGGCTGCCGCTCACTGG 60
 QY 327 LysAsnPhetrAspLysCysTyrTyrrPheSerValcLulysGluIlePheGluAspAla 346
 |||||
 Db 61 AAGAAGCTTCACACAAATGCTACTATTTTCAGTTGAGAAAGAAATTTTGGAGATGCA 120
 QY 347 LysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGluGln 366
 |||||
 Db 121 AAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTCATAAACACTAGAGAGAACAG 180

QY 367 GlnTrpIleLysGlnMetValGlyArgGlyUserHisTrpIleGlyLeuThrAspSer 386
 |||||
 Db 181 CAATGGATAAAAAACAGATGCTAGGAGAGAGAGCCACTGGCTCGCCTCACAGACTCA 240
 QY 387 GluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTrpLysAsnTrpLys 406
 |||||
 Db 241 GAGCGTGAATAATGAATGGAATGGCTGGATGGACATCTCCAGACTACAAAAAATGGAAA 300
 QY 407 AlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeu 426
 |||||
 Db 301 GCTGGACAGCCGGATAACTGGGGTGTATGGCCATGGCCAGAGAGACATGCTGGGTG 360
 QY 427 IleTrpAlaGlyGlnTrpAsnPhsGlnCysGluAspValAsnAsnPhs-IleCysG1 446
 |||||
 Db 361 ATTTATGTCGGCAGGGAACGATTTCCAACTGTAGAGCTCAATAACTTTAATTGGCA 420
 QY 446 uLysAspArgGluThrValLeuSerSerAlaLeu 457
 |||||
 Db 421 AAAAGACAGGAGACAGTACTGGCATCTGCATTA 454
 RESULT 11
 BB248064
 LOCUS
 DEFINITION BB248064 RIKEN full-length enriched, 7 days neonate cerebellum Mus
 musculus cDNA clone A730023E20 3', mRNA sequence.
 ACCESSION BB248064
 VERSION BB248064.2 GI:16355610
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 638)
 Arkawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
 Hiramoto,K., Hori,Y., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
 ,M., Koto,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
 ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
 Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
 Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
 Unpublished (2001)
 On Jul 6, 2000 this sequence version replaced gi:8940810.
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9226
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
 ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
 ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
 ,Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
 ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
 Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
 further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

FEATURES

Location/Qualifiers

source

1. .638
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="A730023E20"
 /clone_lib="RIKEN full-length enriched, 7 days neonate
 cerebellum"
 /tissue_type="cerebellum"
 /dev_stage="7 days neonate"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 20.0 and subtraction to Rot = 459.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5' GAGAGAGAGATTCGAGTAAATAATTAATCCCTCCCTCC
 3']. cDNA was cleaved with XhoI and BamHI. Vector: a
 modified pBluescript KS(+) after bulk excision from Lambda
 FLC I."

BASE COUNT 212 a 134 c 146 g 146 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.7e-47 Length: 638
 Score: 831.00 Matches: 142
 Percent Similarity: 93.25% Conservative: 10
 Best Local Similarity: 87.12% Mismatches: 11
 Query Match: 32.98% Indels: 0
 DB: 10 Gaps: 0

US-09-763-712A-2_COPY_91_547 (1-457) x BB248064 (1-638)

QY 295 ProGlyProGlyProSerGlyAlaValValProLeuAlaLeuGlnAsnGluProThr 314
 |||||
 Db 3 CCTGGCCCTCCAGGCCCTCCAGGAGCAATGGAGCCATGGCTCTGCAGATGAACCAACC 62
 QY 315 ProAlaProGluAspAsnGlyCysProHisTrpLysAsnPheThrAspLysCysTyr 334
 |||||
 Db 63 CCAGCATGAGGATCAACGGATGTCGCCCTCACTGGAAGAACTTCACAGATAAATGCTAC 122
 QY 335 TyrPheSerValGluLysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSer 354
 |||||
 Db 123 TATTTTTCATTTGGAAGAAATTTTGAAGATGCTTAAGCTTTTCTGTGAAGCAAAATCT 182
 QY 355 SerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetVal 374
 |||||
 Db 183 TCCCATCTCGTTTTCATAACTCAAGAGAGAACAGCAATGGATAAAAAAGCATACCGTG 242
 QY 375 GlyArgGluSerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrp 394
 |||||
 Db 243 GGGAGAGAAAGCAATGGATCGCCCTCACAGACTCAGACAGAGAAAGCGAATGGAATGG 302
 QY 395 LeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGly 414
 |||||
 Db 303 CTAGACGGGTCCCTCTTATTACAAAACTGGAAAAGCTGGCAACACAGATAACTGGGGC 362

```

QY 415 HisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnThrPasnAsp 434
Db 363 AGTGGCCATGGGCGGAGGAGAACTGTGCTGTTGTTTACCGACAGAGTGGAAATGAC 422
QY 435 PheGlnCysGluAspValAsnAsnPheIleCysGluLysAspArgGlnThrValLeuSer 454
Db 423 TTCCAGTGTGATGAATCAATAACTTCAATTTGTGAGAGGAAGAGGAGGAGTACCATCA 482
QY 455 SerAlaLeu 457
Db 483 TCCATATTA 491

RESULT 12
BQ934501
LOCUS
DEFINITION BQ934501 884 bp mRNA linear EST 21-AUG-2002
AGENCOURT_8765685 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6330794
5', mRNA sequence.
ACCESSION BQ934501
VERSION BQ934501.1 GI:22349884
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 884)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,
Ph.D.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13785 row: m column: 03
High quality sequence stop: 352.
FEATURES
source
1..884
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6330794"
/clone_lib="NIH_MGC_130"
/lab_host="DH10B (phage-resistant)"
/note="Organ: otocysts; Vector: pCMV-SPORT6.1.cdb;
Site_1: ECKRV; Site_2: NotI; Cloned unidirectionally.
Primer: Oligo dt. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a
NIH_MGC Library."
BASE COUNT 282 a 225 c 205 g 172 t
ORIGIN
Alignment Scores:
Pred. No.: 3,84e-45 Length: 884
Score: 798.50 Matches: 170
Percent Similarity: 90.24% Conservative: 15
Best Local Similarity: 82.93% Mismatches: 12
Query Match: 31.69% Indels: 9
DB: 14 Gaps: 4

US-09-763-712A-2_COPY_91_547 (1-457) x BQ934501 (1-884)
QY 1 MetAsnSerGlnLeuAsnSerPheThr-GlyGlnMetGluAsnIleThrThrIleSerG1 20
Db 277 ATGAATAGCCAGCTACGTCATTTCACAGGTCAGATGGACAACATTACCACCTATCTACA 336
QY 20 nAlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgTh 40
Db 337 GCCACACGACGAGAGCCTGAAGACCTTCAGGACTTACACAGGATACAGAAATAGAAC 396

```

```

QY 40 rAlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValas 60
Db 397 AGCTGTCAAGTTTCAGCAACTTCAGGAACGCTTCCAGGTCTTTGAGACAGATATTGTGAA 456
QY 60 nIleIleSerAsnIleSerTyrThrAlaHisLysLeuArgThrLeuThrSerAsnLeuAs 80
Db 457 CATCATTAGCAACATCAGCTACAGCCCATCACCTGAGGACACTGACCAGCAATCTGAA 516
QY 80 nGluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerIe 100
Db 517 TGATGTTAGGACCACATGCACACACACTTGACCAGACACACCGATGACCTGACCTCCT 576
QY 100 uAsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLe 120
Db 577 GAATAC-ACACTAGTCAACATCCGCTTGGATTCTATTCTCTCAGGATGAGCAAGACAT 635
QY 120 uMetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLy 140
Db 636 GATGAGGTCAAGTTAGACACTGAAGTGCCAACTTATCAGTGGTTATGGAAGAGATGAA 695
QY 140 sLeuValAspSerLysHisGlyGlnLeuLysAsnPheThrIleLeuGlnGlyProPr 160
Db 696 ACTGGTTGACTCAACACCGGTCAGCTCATCAAGAACTTTACCTCTTACAGAGTCTCCTC 755
QY 160 oGlyProArgGly-ProArgGlyAspArgGlySerGlnGlyProPro---GlyProThr- 178
Db 756 TGGCCCCAGAGGTCAAAAAGGGGGACAGAGGATCTCAGGGGACCAACTGGGTCCAACTT 815
QY 179 GlyAsnLysGlyGlnLys-----GlyGluLysGlyGluPro---GlyProPro---Gly 194
Db 816 GGCACAAAGGGGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 875
QY 195 ProAlaGly 197
Db 876 CCTGGCGGC 884

RESULT 13
BQ771366
LOCUS
DEFINITION BQ771366 808 bp mRNA linear EST 26-JUL-2002
IMAGE:5702432 5', mRNA sequence.
ACCESSION BQ771366
VERSION BQ771366.1 GI:21979842
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 808)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
FEATURES
source
Location/Qualifiers
1..808
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5702432"
/clone_lib="NIH_MGC_F10"
/tissue_type="whole brain"
/dev_stage="embryo 12.5dpc"

```

/lab_host="DH10B (T1 phage resistant)"
 /note="Organ: Brain; Vector: pYX-Asc; Site 1: Ecor I;
 Site 2: Not I; The library was constructed according
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with Ecor I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is CAGCCAGC. This library was created for the University
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator."

BASE COUNT 270 a 158 c 173 g 206 t 1 others
 ORIGIN

Alignment Scores:

Pred. No.: 4.92e-43 Length: 808
 Score: 767.00 Matches: 131
 Percent Similarity: 93.33% Conservativity: 9
 Best Local Similarity: 87.33% Mismatches: 10
 Query Match: 30.44% Indels: 0
 DB: 14 Gaps: 0

US-09-763-712A-2_COPY_91_547 (1-457) x BQ771366 (1-808)

Qy 308 AlaleuGlnAsnGluProThrPronlaProGluAspAsnGlyCysProHistrPlys 327
 Db 2 GCTCTGCAGATGACCAACCCAGCATCAGAGTGTCAACGGATGTCGCCCTCACTGGAG 61
 Qy 328 AsnPheThrAspLysCysTyrTyrPheSerValGluLysGluLeuPheGluAspAlaLys 347
 Db 62 AACTTCACAGATAATGCTACTATTTTCATGTGAAAAAGAAATTTTGAAGATGCTAAG 121
 Qy 348 LeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArgGluGlnGln 367
 Db 122 CTTTCTGTGAAGACAAATCTCCATCTCGTTTTCATAACTCAAGAGAACAGCAA 181
 Qy 368 TrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAspSerGlu 387
 Db 182 TGGATAAAAAGCATACCGTGGGGAGAGAAAGCCATTCGATCGGCTCACAGACTCAGAA 241
 Qy 388 ArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAla 407
 Db 242 CAGGAACCGATGGAAGTGTAGACGGTCACTGTGTATTACAAAACCTGGAAAGCT 301
 Qy 408 GlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeu 427
 Db 302 GGACAACCAAGATAACTGGGGCAGTGCCCATGGCCAGGAGAAAGACTGTGCTGGCTGATT 361
 Qy 428 TyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLys 447
 Db 362 TACCGAGGACAGTGAATGACTTCAGTGTGATGAATCAATAACTTCAATTTGTGAGAAG 421
 Qy 448 AspArgGluThrValLeuSerSerAlaLeu 457
 Db 422 GAAAGGAGGAGTACCATCATCATATTA 451

RESULT 14

BM676508/c BM676508 500 bp mRNA linear EST 27-FEB-2002

LOCUS UI-E-EJ0-ahq-c-05-0-UI.s2 UI-E-EJ0 Homo sapiens cDNA clone

DEFINITION UI-E-EJ0-ahq-c-05-0-UI 3', mRNA sequence.

ACCESSION BM676508

VERSION BM676508.1 GI:18986404

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 500)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

source

Location/Qualifiers
 1..500
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-EJ0-ahq-c-05-0-UI"
 /clone_lib="UI-E-EJ0"
 /tissue_type="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina Foveal and Macular, RPE and
 Choroid"
 /dev_stages="fetal and adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: Ecor I; Site 2: Not I;
 UI-E-EJ0 is a subtracted cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an Ecor I adaptor, digested
 with Not I, and cloned directionally into pT7T3-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dr)18 tail. The
 sequence tags for this library are: fetal eyes, AGAATCAAGA
 ; lens, CGATTAGCGA; eye anterior segment, AATGCCGCTAT;
 optic nerve, CCATTAACTG; retina, CCGCG; Retina Foveal and
 Macular, GTCC; RPE and Choroid, ACCTA. This library was
 created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI).
 TAG_LIB=UI-E-EJ0
 TAG_TISSUE=human eye anterior segment
 TAG_SEQ=AATGCCGCTAT"

BASE COUNT 109 a 123 c 89 g 179 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.46e-36 Length: 500
 Score: 669.00 Matches: 117
 Percent Similarity: 100.00% Conservativity: 1
 Best Local Similarity: 99.15% Mismatches: 0
 Query Match: 26.55% Indels: 0
 DB: 14 Gaps: 0

US-09-763-712A-2_COPY_91_547 (1-457) x BM676508 (1-500)

Qy 340 LysGluIlePheGluAspAlaLysLeuPheCysGluAspLysSerHisLeuValPhe 359

Db 495 GAGGAAATTTTGGAGATGCAAGCTTTTCTGTGAGACAAAGTCTTTCACATCTTGTGTTTC 436

Qy 360 IleAsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHis 379

Db 435 ATAAACACATAGAGAGAAACAGCAATGGATGATAAAAAACAGATGGTAGGAGAGAGACCCAC 376

QY 380 TrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSer 399
 |||||
 Db 375 TGGATCGGCTCACAGACTCAGACGCTGAAATGAATGAAGTGGCTGGATGGACATCT 316
 |||||
 QY 400 ProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyPro 419
 |||||
 Db 315 CCAGACTACAAATTTGGAAAGCTGGACAGCCGGATACTGGGGTCATGCCATGGGCCA 256
 |||||
 QY 420 GlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAsp 439
 |||||
 Db 255 GGAGAGACTGCTCGGGTGTATTATGCTGGCAGTGGAGCATTTCCAAATGTGAAGAC 196
 |||||
 QY 440 ValAsnAspPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 |||||
 Db 195 GTCAATAACTTCAATTTGCCGAAAAAGACAGGAGACAGTACTGTCATCTGCATTA 142
 |||||

RESULT 15
 BW713891

LOCUS
 DEFINITION
 UI-E-EJ0-ahq-c-05-0-UI-r1 UI-E-EJ0 Homo sapiens cDNA clone
 UI-E-EJ0-ahq-c-05-0-UI 5', mRNA sequence.

ACCESSION
 BW713891

VERSION
 BW713891.1 GI:19027149

KEYWORDS
 EST.

SOURCE
 human.

ORGANISM
 Homo sapiens

REFERENCE
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 507)

AUTHORS
 Ronaldo,M.F., Lennon,G. and Soares,M.B.

TITLE
 Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL
 Genome Res. 6 (9), 791-806 (1996)

MEDLINE
 9704477

COMMENT
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA
 sequence: 443-498, >Poly_A/simple_repeat

Seq primer: M13 Reverse.
 Location/Qualifiers
 1..507
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-EJ0-ahq-c-05-0-UI"
 /clone_lib="UI-E-EJ0"
 /tissue_type="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina Foveal and Macular, RPE and
 Choroid"
 /dev_stage="fetal and adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-EJ0 is a subtracted cDNA library constructed
 according to Ronaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT73-pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for this library are: fetal eyes, AGAATCAAGA

; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT;
 optic nerve, CCATTAAGTG; retina, CCGCG; Retina Foveal and
 Macular, GTCC; RPE and Choroid, ACGTA. This library was
 created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

BASE COUNT 184 a 89 c 121 g 112 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 3.81e-36 Length: 507
 Score: 663.00 Matches: 116
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 26.31% Indels: 0
 DB: 14 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x BW713891 (1-507)

QY 342 IlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsn 361
 |||||
 Db 3 ATTTTGGAGATGCAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTTGTTCATAAAC 62
 |||||
 QY 362 ThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIle 381
 |||||
 Db 63 ACTAGAGAGGAACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGGATC 122
 |||||
 QY 382 GlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAsp 401
 |||||
 Db 123 GGCCTCAGACTCAGACGGTGGAATAAATGAATGGAAGTGGCTGGATGGACATCTCCAGC 182
 |||||
 QY 402 TyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGlu 421
 |||||
 Db 183 TACAAAAATTTGGAAAGCTGGACAGCCGGATAACTGGGGTCATGGCCATGGGCCAGGAGAA 242
 |||||
 QY 422 AspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsn 441
 |||||
 Db 243 GACTGTCTGGGTGATTATGCTGGGCAGTGGAAACGATTTCCAATGTGAAGACGTCAT 302
 |||||
 QY 442 AsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 |||||
 Db 303 AACTTCATTTGGCAAAAAGACAGGAGACAGTACTGTCATCTGCATTA 350
 |||||

Search completed: March 21, 2003, 12:20:27
 Job time : 1517.54 secs

FEATURES
 source

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 21, 2003, 06:54:47 ; Search time 3021.43 Seconds
(without alignments)
4401.885 Million cell updates/sec

Title: US-09-763-712A-2_COPY_91_547

Perfect score: 2520

Sequence: 1 MNSQLSFTGQENITPISQ.....EDVNNFICEKDKRTVLSSAL 457

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09763712/runat_14032003_100949_18101/app_query.fasta_1.1877
-DB=genEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763712 -ACGN_1_1.6828 @runat_14032003_100949_18101 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREAS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_wa:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_or:*
21: em_ov:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

RESULT 1

ALIGNMENTS

| Result No. | Score | Match % | Length | DB | ID | Description |
|------------|--------|---------|--------|----|-----------|--------------------|
| 1 | 2520 | 100.0 | 2983 | 9 | AB005145 | AB005145 Homo sapi |
| 2 | 2514 | 99.8 | 2641 | 6 | AX047353 | AX047353 Sequence |
| 3 | 2514 | 99.8 | 3058 | 9 | AB038518 | AB038518 Homo sapi |
| 4 | 2487 | 98.7 | 2005 | 6 | AX454442 | AX454442 Sequence |
| 5 | 2487 | 98.7 | 2005 | 6 | AX490920 | AX490920 Sequence |
| 6 | 2344 | 93.0 | 2637 | 10 | AB078434 | AB078434 Mus muscu |
| 7 | 2330 | 92.5 | 3291 | 10 | AB038519 | AB038519 Mus muscu |
| 8 | 2309 | 91.6 | 1886 | 9 | HSM802985 | AL713657 Homo sapi |
| 9 | 2244 | 89.0 | 4330 | 9 | AB052103 | AB052103 Homo sapi |
| 10 | 1021.5 | 40.5 | 130763 | 2 | AC112416 | AC112416 Rattus no |
| 11 | 981 | 38.9 | 182029 | 2 | AC114677 | AC114677 Mus muscu |
| 12 | 981 | 38.9 | 193208 | 2 | AC102618 | AC102618 Mus muscu |
| 13 | 954.5 | 37.9 | 169088 | 2 | AC016128 | AC016128 Homo sapi |
| 14 | 954.5 | 37.9 | 188439 | 9 | AF000915 | AF000915 Homo sapi |
| 15 | 935.5 | 37.1 | 71044 | 2 | AC024368 | AC024368 Homo sapi |
| 16 | 916.5 | 36.4 | 187635 | 2 | AF001022 | AF001022 Homo sapi |
| 17 | 914.5 | 36.3 | 178022 | 2 | AF000900 | AF000900 Homo sapi |
| 18 | 842 | 33.4 | 188255 | 2 | AF000939 | AF000939 Homo sapi |
| 19 | 763 | 30.3 | 188255 | 2 | AF000939 | AF000939 Homo sapi |
| 20 | 747 | 29.6 | 182029 | 2 | AC114677 | AC114677 Mus muscu |
| 21 | 747 | 29.6 | 193208 | 2 | AC102618 | AC102618 Mus muscu |
| 22 | 682 | 27.1 | 3636 | 9 | AB007829 | AB007829 Homo sapi |
| 23 | 682 | 27.1 | 3685 | 6 | E32511 | E32511 Scavenger r |
| 24 | 682 | 27.1 | 3810 | 6 | E32509 | E32509 Scavenger r |
| 25 | 671 | 26.6 | 2215 | 10 | BC026446 | BC026446 Mus muscu |
| 26 | 562 | 22.3 | 188637 | 2 | AL845306 | AL845306 Danio rer |
| 27 | 491.5 | 19.5 | 1265 | 10 | RATSPD | M81231 Rat pulmona |
| 28 | 483.5 | 19.2 | 3508 | 3 | AF053538 | AF053538 Alvinella |
| 29 | 481.5 | 19.1 | 1183 | 10 | BC003705 | BC003705 Mus muscu |
| 30 | 481.5 | 19.1 | 1253 | 10 | MUSSPD | L40156 Mus musculu |
| 31 | 474.5 | 18.8 | 1410 | 6 | AX334792 | AX334792 Sequence |
| 32 | 474.5 | 18.8 | 1410 | 9 | HSMRNAPD | X65018 H.sapiens m |
| 33 | 474 | 18.8 | 1454 | 4 | BTLSPD | X75911 B.taurus mR |
| 34 | 473.5 | 18.8 | 1284 | 4 | AF050950 | AF050950 Bos tauru |
| 35 | 473.5 | 18.8 | 1301 | 9 | BC022318 | BC022318 Homo sapi |
| 36 | 471 | 18.7 | 2377 | 9 | BC008760 | BC008760 Homo sapi |
| 37 | 471 | 18.7 | 5676 | 6 | E07265 | D07265 cDNA encodi |
| 38 | 471 | 18.7 | 5676 | 9 | HUMCA1V | D90279 Human mRNa |
| 39 | 471 | 18.7 | 7138 | 9 | HUMPA1V | M76729 Human pro-a |
| 40 | 465 | 18.5 | 6114 | 10 | CRUPA1V | M76730 Chinese ham |
| 41 | 462.5 | 18.4 | 4990 | 10 | CRUCOLVIA | L06863 Cricetulus |
| 42 | 462 | 18.3 | 2777 | 5 | GGU07973 | U07973 Gallus gall |
| 43 | 462 | 18.3 | 4428 | 6 | AX146422 | AX146422 Sequence |
| 44 | 462 | 18.3 | 4428 | 6 | AX146424 | AX146424 Sequence |
| 45 | 462 | 18.3 | 5575 | 5 | AF137273 | AF137273 Gallus ga |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

```

AB005145
LOCUS       AB005145       2983 bp    mRNA    linear    PRI 21-NOV-2001
DEFINITION Homo sapiens CL-P1 mRNA for collectin placenta 1, complete cds.
ACCESSION  AB005145
VERSION    AB005145.1  GI:17026100
KEYWORDS   .
SOURCE     Homo sapiens female tissue_lib:placenta cdna to mRNA.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
AUTHORS    Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Keshi,H.,
            Sakai,Y., Fukuh,A., Sakamoto,T., Itabe,H., Suzutani,T.,
            Ogasawara,M., Yoshida,I. and Wakamiya,N.
TITLE      The membrane-type collectin CL-P1 is a scavenger receptor on
            vascular endothelial cells
JOURNAL    J. Biol. Chem. 276 (47), 44222-44228 (2001)
MEDLINE    21570232
REFERENCE  2 (bases 1 to 2983)
AUTHORS    Ohtani,K.
TITLE      Direct Submission
JOURNAL    Submitted (25-JUN-1997) Katsuki Ohtani, Asahikawa Medical College,
            Department of Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa,
            Hokkaido 078-8510, Japan (E-mail:ohtani@asahikawa-med.ac.jp,
            Tel:+81-166-68-2393, Fax:+81-166-68-2399)
FEATURES   Location/Qualifiers
            1..2983
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /sex="female"
             /tissue_lib="placenta"
            1..2983
             /gene="CL-P1"
             /gene="CL-P1"
             /codon_start=1
             /product="collectin placenta 1"
             /protein_id="BAB72147.1"
             /db_xref="GI:17026101"
             /translation="MKDDFAEEVEVSFGYKRFQIGTQCTKCKNNWALKPSILLY
            ILCALLTIVAILGVKVKMDNTVGMETSROTVDDKLLTAVESDLKLGDDTGKKA
            STNSLSTFRSDILDLROQLREITEKTSKNDTLEKLAQSGDALVDRSOLKETLENN
            SFLITVNTKLTQAYNGYVTLNQDDTSVLQGNLQNMYSHVVIMMNLNLTFVQQRN
            LITNQRSVDDTSQAIRIKNDQFNLOQVFOAKKDTDLKXKVUSLQTLAANSALA
            KANDTLDMSLQSFQOMENITTSQANEQNLKLDQLDKDAENETAIKFNQLEE
            RFLPETDIVNISVTAHLRLTSLNNEVRTCTDLTKHDDDLTSLNLTANI
            RLDSVLRMQDLMLRSRLDTEVANLSVIMEKMLVDSKHGOLIKNFTILOGPPGPRGP
            RGDRSQPPGPTGNKGQKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEKGEK
            GSRSPGKPGQPGSPGDPGPPGKGLPGQPGPPGQFQGLQGTGVEGVPGPGLPG
            LPVFGPGKPGPPGPGSGAVPALQNEPTPAPEDNGCPPHKNFTDKCYEVSVE
            KEIFEDAKLFCEDKSSHLVFINTREEQWIKQWVGRESHWIGLWDSRENEKWLKG
            TSPDYKNNKAGQPDNMWGHGHPGEDCAGLIYAGQWDFQCEDVNNFICEKDRETVLSS
            AL"
BASE COUNT 914 a 707 c 703 g 659 t
ORIGIN
Alignment Scores:
Pred. No.:      2,57e-82      Length:      2983
Score:          2520.00      Matches:     457
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     100.00%      Indels:      0
DB:              9           Gaps:         0

US-09-763-712A-2_COPY_91_547 (1-457) x AB005145 (1-2983)
QY      1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
        |||||||
DB      926 ATGAACAGCAGCTCAACTCATTACAGGTCAGATGGAGAACATCACCATTCTCTCAA 985
        |||||||
QY      21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuUHisLysAspAlaGluAsnArgThr 40
        |||||||
DB      986 GCAACGAGCAGAACCTGAAGACCTGCAGGACTTACACAAAGATGCAGAGATAGAACA 1045
        |||||||

```

```

QY      41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
        |||||||
DB      1046 GCCATCAAGTTCAACAACACTGGAGGAACGCTTCCAGCTCTTTGAGACGGATATTGTGAAC 1105
        |||||||
QY      61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrIleUHisSerAsnLeuAsn 80
        |||||||
DB      1106 ATCATTTAGCAATATCAGTTTACACAGCCACCCTCGGACGCTGACCAATCTAAAT 1165
        |||||||
QY      81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
        |||||||
DB      1166 GAAGTCAGAGACCACTTGCACAGATACCTTACCAACACACAGATGATCTGACCTCTCTG 1225
        |||||||
QY      101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
        |||||||
DB      1226 AATAATACCTCGCCCAACATCCGTTTGGATTCTGTTCTCTCAGGATGCAACAAGATTG 1285
        |||||||
QY      121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
        |||||||
DB      1286 ATGAGTCGAGGTTAGACACTGAAGTAGCCAATTTATCAGTGATTATGAAGAAATGAAG 1345
        |||||||
QY      141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
        |||||||
DB      1346 CTAGTAGACTCCAAGCATGCTCAGCTCATCAAGATTTTACAATACTACAAGTCCACCG 1405
        |||||||
QY      161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGlyAsn 180
        |||||||
DB      1406 GGCCCCAGGGGTCCAAGAGGTCACAGAGGATCCAGGGACCCCTGGCCCACTGGCAAC 1465
        |||||||
QY      181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGluArgGly 200
        |||||||
DB      1466 AAGGGACAGAAAGGAGAGAGGGGAGCCTGGACACCTGGCCCTCGGGGTGAGAGGC 1525
        |||||||
QY      201 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
        |||||||
DB      1526 CCNATTTGGACAGCTGGTCCCCCGGAGAGCGCTGGCGCAAGAGATCTAAAGCTCCAG 1585
        |||||||
QY      221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlyProGlyProSerGlyAsp 240
        |||||||
DB      1586 GGCCCCAAAGGGTCCCGTGGTTCCTGGGAAGCCCGGCCCTCAGGGCCCTCAGTGGGAC 1645
        |||||||
QY      241 ProGlyProGlyProGlyProGlyLysGlyGluGlyLeuProGlyProGlyProGlyPro 260
        |||||||
DB      1646 CCAGGCCCCCGGGGCCACCCAGGAAAGAGGACTCCCCGGGCCCTCAGGGCCCTCTCGC 1705
        |||||||
QY      261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
        |||||||
DB      1706 TTCCAGGGACTTCAGGGCACCGTTGGGAGCCTGGGGTGCCTGGACCTCGGGGACTGCCA 1765
        |||||||
QY      281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
        |||||||
DB      1766 GGCTTGCCTGGGTACAGGCATGCCAGGCATGCCAGGCCCCCAAGGGCCCCCGGCCCTCTCGGCCCA 1825
        |||||||
QY      301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
        |||||||
DB      1826 TCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAAATGACCAACCCCGCACGGAGGACAAT 1885
        |||||||
QY      321 GlyCysProProHisThrLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
        |||||||
DB      1886 GGCTGCCCGCCTCAGCTGGAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1945
        |||||||
QY      341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
        |||||||
DB      1946 GAAATTTTGGAGGATGCAAGGCTTTTCTGTGAAGACAAAGCTTTCACATCTTGTGTTTTCATA 2005
        |||||||
QY      361 AsnThrArgGluGlnGlnGlnThrPheLysLysGlnMetValGlyArgGluSerHisTrp 380
        |||||||
DB      2006 AACACTAGAGAGAACACCAATGGATAAAAAACAGATGGTAGGGAGAGAGAGCCACTGG 2065
        |||||||
QY      381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
        |||||||
DB      2066 ATCGGCCTCACAGACTCAGAGCGGTGAAGATGAATGAAGTGGCTGGATGGGACATCTCCA 2125
        |||||||

```


QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
|||||
Db 2126 GACTACAAAATTGGAAGCTGGACAGCCGATAAAGTGGGTCATGGCCATGGGCCAGGA 2185
QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
|||||
Db 2186 GAAGACTGCTGGGGTTGATTTATGCTGGCAGTGAAGCATTTCCAATGTGAAGACGTC 2245
QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
|||||
Db 2246 AATACTTCAATTTGCGAAAAGACAGAGGAGACAGTACTGTCTCATTA 2296
RESULT 2
AX047353 2641 bp DNA linear PAT 15-DEC-2000
LOCUS
DEFINITION
Sequence 39 from Patent WO0068380.
ACCESSION
AX047353
VERSION
AX047353.1 GI:11876599
KEYWORDS
human.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 2641)
AUTHORS
Bandman,O., Hillman,J.L., Tang,Y.T., Lal,P., Yue,H., Baughn,M.R.,
Lu,D.A. and Azimzai,Y.
TITLE
Extracellular matrix and adhesion-associated proteins
JOURNAL
Patent: WO 0068380-A 39.16-NOV-2000;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
source
1..2641
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 4586187CB1"
BASE COUNT 823 a 628 c 608 g 582 t
ORIGIN
Alignment Scores:
Pred. No.: 3.82e-82 Length: 2641
Score: 2514.00 Matches: 456
Percent Similarity: 99.78% Conservative: 0
Best Local Similarity: 99.78% Mismatches: 1
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0
US-09-763-712a-2_copy_91_547 (1-457) x AX047353 (1-2641)
QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
Db 567 ATGAACAGCGCAGCTCAACTCACTTACAGGTCAGATGGAGAATCACTATCTCTCAA 626
QY 21 AlaAspGluGlnAsnLeuLysAspLeuGlnAspLeuHisIlyAspAlaGluAsnArgThr 40
Db 627 GCCAACGACAGACCTGAAAGACCTGACAGGACTTACAAAGATGACAGAAATAGAACA 686
QY 41 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 60
Db 687 GCCATCAAGTTCAACCACTGGAGGAACGCTTCCAGCTCTTGGAGACGATATTGTGAAC 746
QY 61 IleIleSerAsnIleSerThrAlaHisHisLeuAspThrLeuThrSerAsnLeuAsn 80
Db 747 ATCATTTAGCAATATCATAGTTACAGCCACCACCTCCGACGCTGACGACGATCTAAAT 806
QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
Db 807 GAAGTCAGGACCACTTGCACAGATACCCCTTACCAACACACAGATGATCTGACCTCTTG 866
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
Db 867 AATAATACCTGGCCAACTCCGTTGGATTCTGTTCTCTCAGGATGCAACAAGATTG 926
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluMetLys 140
|||||

Db 927 ATGAGTCGAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAAGAAATGAAG 986
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
|||||
Db 987 CTAGTAGACTCCAGGATGGTCACTCCTCAAGAAATTTTACAATACTACAAAGTCCACCG 1046
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGlyAsn 180
|||||
Db 1047 GGGCCCGGGTCCAAAGAGGTGACAGAGATCCAGGAGCCCTGGCCCACTGGCCAACTGGCAAC 1106
QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGlyGluArgGly 200
|||||
Db 1107 AAGGGACAGAAAGAGAGAGAGGGGAGCGCTGGACCACTGGCCCTGCGGTGAGAGAGGC 1166
QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
|||||
Db 1167 CCATTTGGACCACTGGTCCCCCGGAGAGCGTGGCGGCAAGATCTTAAAGGCTCCAG 1226
QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
|||||
Db 1227 GGGCCCAAGGCTCCCGTGGTCCCTGGGAAGCCCGCCCTCAGGGCCCGCAGTGGGGAC 1286
QY 241 ProGlyProProGlyProGlyLysGlyGluLeuProGlyProGlnGlyProGlyPro 260
Db 1287 CCAGGCCCCCGGGCCACACAGGCAAGAGGAGACTCCCGGGCCCTCAGGGCCCTCTCTGGC 1346
QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
Db 1347 TTCAGGGGACTTCAGGGCACCGTTGGGGAGCGTGGGGTGGCTGGACCTCGGGGACTGCCA 1406
QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
Db 1407 GGCTTGCCTGGGGTACCAGGCATGCCAGGCCCAAGGGCCCGCCCTCTCTGGCCCA 1466
QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
Db 1467 TCAGGAGCGGTGGTGGCCCTGGCCCTGCAGAATGAGCAACCCCGCCAGGAGCAAT 1526
QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
Db 1527 AGCTGGCCGCTCCTCCTGGGAGAACTTCAACAGCAAAATGCTACTATTTTTCAGTTGAGAA 1586
QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
Db 1587 GAAATTTTTCAGGATGCAAAAGCTTTTCTGTGAAGACAAGTCTTACATCTTGTTCATA 1646
QY 361 AsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
Db 1647 AACACTAGAGAGAGACAGCAATGGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGG 1706
QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
Db 1707 ATCGGCTCACAGACTCAGAGCGTGAATGAATGAATGGTGGTGGTGGGACATCTCCA 1766
QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
Db 1767 GACTACAAAAATTGGAAGCTGGACAGCCGATAAAGTGGGGTCACTGGCCATGGGCCAGGA 1826
QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
Db 1827 GAAGACTGCTGGGTGATTTATGCTGGCAGTGGCAACGATTTCCAATGTGAAGACGTC 1886
QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
Db 1887 AATACTTCAATTTGCGAAAAGACAGGAGAGAGTACTGTCTCATTA 1937
RESULT 3
AB038518 3058 bp mRNA linear PRI 08-MAR-2001
LOCUS
DEFINITION
Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
type 1, complete cds.
ACCESSION
AB038518
VERSION
AB038518.1 GI:13365514

QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 |||||
 Db 2324 GAAGACTGTGCTGGGTTGATTTATGCTGGCAGTGGAAAGGATTTCAATGTTGAAGACGTC 2383
 |||||
 QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 |||||
 Db 2384 AATAACTTCATTGGAAAAAGACAGGGAGACAGTACTGTCTATCTGCATTA 2434
 |||||
 RESULT 4
 AX454442
 LOCUS AX454442 2005 bp DNA linear PAT 06-JUL-2002
 DEFINITION Sequence 27 from Patent WO0208284.
 ACCESSION AX454442
 VERSION AX454442.1 GI:21713845
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1
 Baker, K.P., Ferrara, N., Gerber, H., Gerritsen, M.E., Goddard, A.,
 Godowski, P.J., Gurney, A.L., Hillan, K.J., Marsters, S.A., Pan, J.,
 Paoni, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I.
 and Ye, W.
 TITLE Compositions and methods for the diagnosis and treatment of
 disorders involving angiogenesis
 JOURNAL Patient: WO 0208284-A 27 31-JAN-2002;
 Genentech, Inc. (US) ; Baker, Kevin P. (US) ; Ferrara, Napoleone
 (US) ; Gerber, Hanspeter (US) ; Gerritsen, Mary E. (US) ; Goddard,
 Audrey (US) ; Godowski, Paul J. (US) ; Gurney, Austin L. (US) ;
 Hillan, Kenneth J. (US) ; Marsters, Scot A. (US) ; Pan, James (US)
 ; Paoni, Nicholas F. (US) ; Stephan, Jean-Philippe F. (US) ;
 Watanabe, Colin K. (US) ; Williams, P. Mickey (US) ; Wood, William
 I. (US)
 FEATURES
 Location/Qualifiers
 1..2005
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 606 a 493 c 491 g 415 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,84e-81 Length: 2005
 Score: 2487.00 Matches: 455
 Percent Similarity: 99.56% Conservative: 0
 Best Local Similarity: 99.56% Mismatches: 2
 Query Match: 98.69% Indels: 1
 DB: 6 Gaps: 0
 US-09-763-712A-2_COPY_91_547 (1-457) x AX454442 (1-2005)
 QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
 |||||
 Db 134 ATGAACAGCCACTCACTCATTCAGGTGAGTGGAGAACATCACCCTATCTCTCAA 193
 |||||
 QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
 |||||
 Db 194 GCCAACGAGCAACCTGAAAGACCTGCAGGACTTACACAAAGATGCAGAGAATAGAACA 253
 |||||
 QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
 |||||
 Db 254 GCCATCAAGTTCACCAACTGGAGGAACGCTTCCAGCTCTTTGAGACGGATATTGTGAAC 313
 |||||
 QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
 |||||
 Db 314 ATCATTAGCAATATCAGTTACACAGCCACCCTGCGGACCGCTGACCAGCAATCTAAAT 373
 |||||
 QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
 |||||
 Db 374 GAAGTCAGGACCACCTGCACAGATACCTT-ACCAACACACAGATGATCTGACCTCTTG 432
 |||||
 QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
 |||||

Db 433 AATAATACCCCTGGCCAAACATCCGTTTGGATCTGTTCTCTCAGGATGCAACAAGATTGG 492
 |||||
 QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
 |||||
 Db 493 ATGAGGTCGAGGTTAGACACTGAAGTAGCAACTTATCAGTGTATGATGAAAGAAATGAAG 552
 |||||
 QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
 |||||
 Db 553 CTAGTAGACTCCCAAGCATGGTCAGCTCATCAAGAATTTTACAATACTACAAGGTCCACGG 612
 |||||
 QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGlyAsn 180
 |||||
 Db 613 GGGCCCGAGGGTCCCAAGAGGTGACAGAGATCCCAAGGACCCCTGGCCCACTGGCAAC 672
 |||||
 QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyProAlaGlyGluArgGly 200
 |||||
 Db 673 AAGGACAGAAAGGAGAGAGGGGGAGCTGGACCACCTGGCCCTGCGGGTGAGAGAGGC 732
 |||||
 QY 201 ProIleGlyProAlaGlyProProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 |||||
 Db 733 CCAATTGGACCACTGGTCCCGCCGAGAGCGTGGCGCAAGGATCTTAAGGCTCCCAAG 792
 |||||
 QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
 |||||
 Db 793 GGGCCCAAGGCTCCCGTGGTTCCCTGGGAAGCGCGCCCTCAGGGCCCCAGTGGGGAGC 852
 |||||
 QY 241 ProGlyProProGlyProProGlyLysGlyGluGlyLeuProGlyProGlnGlyProGly 260
 |||||
 Db 853 CCAGGCCCCCGGGCCACCAGCAAGAGGAGACTCCCGCGCCCTCAGGGCCCTCTCTGGC 912
 |||||
 QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 |||||
 Db 913 TTCCAGGAGACTTCAGGGCCACCGTTGGGAGCGCTGGGTGCCTGGACTCGGGGACTGCCA 972
 |||||
 QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProProGlyPro 300
 |||||
 Db 973 GGGCTGGCTGGGGTACCAGGATGCCAGGCCCAAGGGCCCCCGGGCCCTCTCTGGCCCA 1032
 |||||
 QY 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 |||||
 Db 1033 TCAGGAGCGGTGGTGGCCCTGGCCCTGCAGATGAGCAACCCCGCCGAGGAGACAAT 1092
 |||||
 QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTrpPheSerValGluLys 340
 |||||
 Db 1093 AGCTGCCCGCCCTCACTGGAGAACTTTCACAGACAAATGCTACTATTTTTCAGTTGAGAAA 1152
 |||||
 QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
 |||||
 Db 1153 GAAATTTTGGAGGTGCAAGCTTTTCTGTGAAGACAAGTCTTCACATCTTGTTCATA 1212
 |||||
 QY 361 AsnThrArgGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 |||||
 Db 1213 AACACTAGAGAGAAACAGCAATGGATGATAAAAAACAGATGGTAGGAGAGAGAGCCACTGS 1272
 |||||
 QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 |||||
 Db 1273 ATCGGCTCACAGACTCAGAGCGTGAATAATGAATGAAGTGGCTGGTGGACATCTCCA 1332
 |||||
 QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyProGly 420
 |||||
 Db 1333 GACTCAAAAATTTGAAAGCTGACAGCGGATAACTGGGGTCATGGCCATGGCCGAGGA 1392
 |||||
 QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 |||||
 Db 1393 GAAGACTGTGCTGGGTTGATTTATGCTGGCAGTGGAAAGATTTCCAATCTGAAGAGCTC 1452
 |||||
 QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 |||||
 Db 1453 AATAACTTCATTGGAAAAAGACAGGGAGACAGTACTGCTCTGCAATTA 1503
 |||||
 RESULT 5
 AX490920
 LOCUS AX490920 2005 bp DNA linear PAT 16-AUG-2002

DEFINITION Sequence 27 from Patent WO0200690.
ACCESSION AX490920
VERSION AX490920.1 GI:22323797
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Baker, K.P., Ferrara, N., Gerber, H., Gerritsen, M.E., Goddard, A.,
Godowski, P.J., Gurney, A.L., Hillan, K.J., Marsters, S.A., Pan, J.,
Paoni, N.F., Stephan, J.P., Watanabe, C.K., Williams, P.M., Wood, W.I.
and Ye, W.
TITLE Compositions and methods for the diagnosis and treatment of
disorders involving angiogenesis
JOURNAL Patent: WO 0200690-A 27 03-JAN-2002;
Genentech, Inc. (US)
FEATURES
source Location/Qualifiers
1..2005
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 606 a 493 c 491 g 415 t
ORIGIN
Alignment Scores:
Pred. No.: 2.84e-81 Length: 2005
Score: 2487.00 Matches: 455
Percent Similarity: 99.56% Conservative: 0
Best Local Similarity: 99.56% Mismatches: 2
Query Match: 98.69% Indels: 1
DB: 6 Gaps: 0
US-09-763-712a-2_copy_91_547 (1-457) x AX490920 (1-2005)
QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
DB 134 ATGAACAGCCAGCTCAACTCACTACAGGTGAGTGGAGAACATCACCCTCTCTCAA 193
QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
DB 194 GCCAACGACGACAACTGAAGACCTGCAGACTTACACAAAGATGCAGAGAATAGAAC 253
QY 41 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 60
DB 254 GCCATCAAGTTCAACCAACTGAGGAGAACCTTCCAGCTCTTGACAGCGGATATGTGAAC 313
QY 61 IleIleSerAsnIleSerThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
DB 314 ATCATTAGCAATATCAGTTACAGACCCACCTCCAGCTGCGGACGCTGACCAATCTAAAT 373
QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
DB 374 GAAGTCAGACCCACTGTCACAGATACCTT-ACCAACACACAGATGATCTGACCTCCTTG 432
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
DB 433 AATAATACCCTGCCCAACATCCGTTGGATTCTGTTCTCTCAGGATGCAACAAGATTG 492
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
DB 493 ATGAGGTCGAGGTTAGACACTGAAGTAGCCAACTTATCATGATTATGGAAGAAATGAAG 552
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
DB 553 CTAGTAGACTCCAAGCATGTCAGCTCATCAAGATTTTACAACTACTACAAGGTCACCG 612
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProGlyAsn 180
DB 613 GGCCCCAGGGGTCCAAGAGGTGACAGAGGATCCAGGGGACCCCTGGCCCAACTGGCAAC 672
QY 181 LysGlyGlnLysGlyGlyGluProGlyProGlyProGlyProGlyProGlyGluArgGly 200
DB 673 AAGGGACAGAAAGGAGAGAGGGGAGCGCTGGACCACTGGGCCCTCGGGGTGAGAGAGGC 732

QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
DB 733 CCAATTTGGACAGCTGGTCCCTCCCGGAGAGCGTGGCGGCAAGGATCTAAAGGCTCCAG 792
QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
DB 793 GGCCCCAAAGGCTCCCGTGGTTCCCTGGGAAAGCCGCGCCTCAGGGCCCAAGTGGGGAC 852
QY 241 ProGlyProGlyProGlyProGlyLysGlyGluLeuProGlyProGlnGlyProProGly 260
DB 853 CCAGGCCCCCGGGGCCACACAGGAAAGAGGACTCCCGGCCCTCAGGGCCCTCCTGGC 912
QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
DB 913 TTTCAGGGACTTCAGGGCACCGTTGGGAGACCTGGGGTGCCTGGACCTCGGGACTGCCA 972
QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
DB 973 GGCTTGCCTGGGTACACAGGCATGCCAGGCCCAAGGGCCCGCCCGCTCCTGGCCCA 1032
QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
DB 1033 TCAGGAGCGGTGGTCCCTGCGCTGCAGATGAGCCCAACCCCGCACCGGAGGACAAT 1092
QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
DB 1093 AGCTGCCCGCCCTCACTGGGAAGAACTTCACAGACAAATGCTACTATTTTCAGTTGAGAAA 1152
QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIle 360
DB 1153 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAGCTTTCACATCTGTGTTTCATA 1212
QY 361 AsnThrArgGluGlnGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
DB 1213 AACACTAGAGAGAACAGCAATGGATAAAAAACAGATGTTAGGAGAGAGAGCCACTGG 1272
QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
DB 1273 ATCGGCTCACAGACTACAGCGTGAATAAATGAATGAAGTGGTGGATGGGACATCTCCA 1332
QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
DB 1333 GACTACAAAAATTTGAAAGCTGGACAGCCGGATAACTGGGGTCATGGCCATGGCCACAGA 1392
QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
DB 1393 GAAGACTGTGCTGGGTGATTATGCTGGGCAGTGGAAACGATTTCCAAATGTGAGACGTC 1452
QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
DB 1453 AATAACTTCATTTGGAAAAAGACAGGAGACAGTACTGTCATCTGCATTA 1503

RESULT 6
AB078434
LOCUS AB078434
DEFINITION Mus musculus CL-P1 mRNA for collectin placenta 1, complete cds.
ACCESSION AB078434
VERSION AB078434.1 GI:21901968
KEYWORDS
SOURCE Mus musculus tissue_lib:Liver cdna to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
AUTHORS Ohtani, K., Suzuki, Y., Eba, S., Kawai, T., Kase, T., Keshi, H.,
Sakai, Y., Fukuo, A., Sakamoto, T., Itabe, H., Suzutani, T.,
Ogasawara, M., Yoshida, I., and Wakamiya, N.
TITLE cDNA cloning of mouse CL-P1 gene
JOURNAL Unpublished
REFERENCE
AUTHORS Ohtani, K., Suzuki, Y., Eba, S., Kawai, T., Kase, T., Keshi, H.,
Sakai, Y., Fukuo, A., Sakamoto, T. and Wakamiya, N.

TITLE Direct Submission
JOURNAL Submitted (18-JAN-2002) Katsuki Ohtani, Asahikawa Medical College, Microbiology; 2-1-1 Midorigaoka-Higashi, Asahikawa, Hokkaido 078-8510, Japan (E-mail: ohtani@asahikawa-med.ac.jp, Tel: 81-166-68-2393, Fax: 81-166-68-2399)

FEATURES
source

1. .2637
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /tissue_lib="Liver"

gene

1. .2637
 /gene="CL-P1"

CDS

92. .2320
 /gene="CL-P1"
 /codon_start=1
 /product="collectin placenta 1"
 /protein_id="BAC05523.1"
 /db_xref="GI:21901969"

/translation="MKDDFAEEVEVQFGYKRFQIQGTQCTCKNNWALKFSLVLLY
 ILCALLITVAIIQYKVMKNDVTDGMTSHQTYDNKLTAVESDLKKLGDQAGKAL
 STNSELSTFRSDILDRLQQLQEIETKSKNKDLEKLEQANGDSLVDQSOLKETLQNN
 SFLITTNKTKLQAVNGYVTLNQDTSVLQNLQSOYTSQSVIMNMLNLTQVOORNN
 LINSNQQSDVDSIAQRIKNDQNFQLOVFOAKKDDDLKKEVQSLOTAAANNSALA
 KANNDTLEDANSOLSSFTGQMDNITTTISQANEQSLKDLQDLKDTERTAVKFSQLEE
 RFQVFEDIVNLSNISTAYAHLLATLSLNQNDVRICTDILTRHTDLDLTSLNNTLVNI
 RLDSISLRQMDMMRSKLDTEVANLSVVMEEMLKLVDSKHGLIKNFTILOQPPGPRGP
 KGDGSGQPDPTGNKQKGEKGPDPGPPGAGERTIGPVPPGERSKSGSKSGQPK
 GSRGSPKPGQSPGSPGPPGPPGKDLPGQPPGPGFQGLQGTGVPGPVGPGRGLPG
 LPQVPMGPGKPGPPGPGSGAMEPLALONEPTPASEVNCSPPHKWNFTDKCYVFSLE
 KEIPEADKLCEDEKSSHLVFINREEQOWIKHTVGRSHWIGLITSEQSEKWLQDG
 SPVDFKNNKAGQPDNWSGSHGPGEDCAGLIYAGQWDFQCDENINFCEREAAPSS
 IL"

BASE COUNT 829 a 617 c 635 g 556 t

ORIGIN

Alignment Scores:

Pred. No.: 4.44e-76 **Length:** 2637
Score: 2344.00 **Matches:** 417
Percent Similarity: 96.50% **Conservative:** 24
Best Local Similarity: 91.25% **Mismatches:** 16
Query Match: 93.02% **Indels:** 0
DB: 10 **Gaps:** 0

US-09-763-712A-2_COPY_91_547 (1-457) x AB078434 (1-2637)

Qy 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
 Db 947 ATGATACCGAGCTCAGCTCATTCACAGTTCAGTGGACATTAACCACTATTCACAG 1006
 Qy 21 AlaAsnGluGlnAsnLeuAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
 Db 1007 GCCAACGAGCAGAGCTCAAGACCTTCAGGACTTACACAGGATACAGAAATAGACA 1066
 Qy 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
 Db 1067 GCTGTCAAGTTTCAGCAACTTGGAGAGCGCTTCCAGGCTTTGAGACAGATATTGTGAAC 1126
 Qy 61 IleIleSerAsnIleSerThrAlaHisLysLeuArgThrLeuThrSerAsnLeuAsn 80
 Db 1127 ATCATTCACATCAGCTACACGCCCATCCCTGGAGACACTGACCAGCAATCTGAAT 1186
 Qy 81 GluValArgThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
 Db 1187 GATGTTAGACCACATGCACAGACACCTTGACCAGACACAGGATGACCTGACCTCTTG 1246
 Qy 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnAspLeu 120
 Db 1247 AATAACACACTAGTCAACATCGCTTGGATTCATTCTCTCAGGATGCAGCAAGACATG 1306
 Qy 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
 Db 1307 ATGAGGTCAAAGTTAGACACTGAAGTGGCCCACTTATCAGTGGTTATGGAACAGATGAA 1366

Qy 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
 Db 1367 CTGGTTGACTCAACACCGGTGAGCTCATCAAGAACTTTTACCATTCTACAAGGTCTCTCT 1426
 Qy 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProProGlyProThrGlyAsn 180
 Db 1427 GGCCCCAGAGGTCCAAAGGTGACAGAGATCTCAGGAGCACCTGGTCCAACTGGCAAC 1486
 Qy 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProProGlyProAlaGlyGluArgGly 200
 Db 1487 AAGGACAGAAAGAGAGAGAGAGAGAGCTGTGTCCACCTGGCCCTCGGGGTGAGAGGGC 1546
 Qy 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 Db 1547 ACAATTGGACCACTCGGCCCTCTGTGAGAGCGTGCAGCAAGAGATCCAAAGGCTCACAG 1606
 Qy 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
 Db 1607 GGTCCCAAGAGATCTGTGGGTGCCAGGAAGCTGGCCCTCAAGGACCTAGTGGGGAC 1666
 Qy 241 ProGlyProProGlyProGlyLysGluGlyLeuProGlyProGlnGlyProProGly 260
 Db 1667 CCAGGACCAACAGGTCCACAGGCAAGGATGACTCCCTGGCCCTCAGGGCCCTCTCTGGC 1726
 Qy 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 Db 1727 TTCAGGGACTACAGGGCACTGTGGGTGAGCTGAGTACCTGGAGCTCGGGGGTTGCCA 1786
 Qy 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProProGlyProProGlyPro 300
 Db 1787 GCTTGGCAGGGGTGCCAGGCACTGTGGGTGAGCTGAGGACCACTGGCCCTCCAGGGCCCC 1846
 Qy 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 Db 1847 TCAGGAGCAATGGAGCCATTCGCTCTGCAGATGAACACACCCAGCATCAGAGTCAAC 1906
 Qy 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
 Db 1907 GGATGTCGCCCTCACTGGGAAGAACTTCACAGATAAATGCTACTATTTTCATTGGAAAA 1966
 Qy 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerHisLysLeuValPheIle 360
 Db 1967 GAAATTTTGAAGATGCTAAGCTTTTCTGTGAAGACAAATCTCCCATCTCGCTTTTCATA 2026
 Qy 361 AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 Db 2027 AACTCAAGAGAGACACACCAATGGATAAAAAAGCATACCGTGGGAGAGAAACCATGG 2086
 Qy 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 Db 2087 ATCGCCCTCACAGACTCAGAACAGGAAAGCAATGGAAGTGGCTAGACGGGTCACTGTT 2146
 Qy 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
 Db 2147 GATTACAAAACTGGAAGCTGGACACACAGATAACTGGGCGAGTGGCCATGGGCACGGA 2206
 Qy 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 Db 2207 GAAGACTGTGTGGCTTGATTTACCGACAGACTGGAATGACTTCCAGTGTGATGAATC 2266
 Qy 441 AsnAsnPheIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
 Db 2267 AATACTTCATTGTGAGAGAAAGGAGGAGGAGGAGGAGTACCATCATCATATTA 2317

RESULT 7

AB038519

LOCUS

DEFINITION

AB038519

VERSION

KEYWORDS

SOURCE

3291 bp mRNA linear ROD 18-JAN-2002
 Mus musculus src1 mRNA for scavenger receptor with C-type lectin,
 complete cds.
 AB038519 GI:18146951
 Mus musculus cdna to mRNA.

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
 AUTHORS Nakamura, K., Funakoshi, H., Tokunaga, F. and Nakamura, T.
 TITLE Molecular cloning of a mouse scavenger receptor with C-type lectin (SRCL)(1), a novel member of the scavenger receptor family
 JOURNAL Biochim. Biophys. Acta 1522 (1), 53-58 (2001)
 MEDLINE 21575692

REFERENCE 2 (bases 1 to 3291)
 AUTHORS Nakamura, K. and Nakamura, T.
 TITLE Direct Submission
 JOURNAL Submitted (18-FEB-2000) Kenji Nakamura, Osaka University Medical School, Division of Biochemistry, Biomedical Research Center; 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
 (E-mail: knakamura@onbiomed.osaka-u.ac.jp,
 Tel.:81-6-6879-3783(ex.3783), Fax:81-6-6879-3789)

FEATURES
 source 1..3291
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 gene 1..3291
 /gene="srcl"
 CDS 77..2305
 /gene="srcl"
 /codon_start=1
 /product="scavenger receptor with C-type lectin"
 /protein_id="BAB82497.1"
 /db_xref="GI:1814692"
 /translation="MKDDFAEEVEQSFGRFGIHEGTQCTKCNWALKFSIVLLYILCALLTIVAILGYKVKVMDNSVDMETSHQTYDNKLTAVESDLKLGDAQKKALSTNSELSTFIDLRQLOEITEKTSKNDTLEKLAQNGDSLVDROSLKETLQNNSLFTTNTKTLQAVNGYVNLQDNTNVLQNLQSQSVYINNNLNLTVQOORNLTSLNQSVDDTSLAIORIKNDFONLQOVFLQAKDQDLKQVSLQTLAANNSALAKANDTLEDMSQLSFTGQMDNLTTSQANEQSLKDLQDLHKDTENTAYKFSOLEE RFQFETDIWNIISNISHTAHLRFLTSLNNDVWTCTDTRHTRDDTLNNTLVNIRLDSLSLRQMKRSKLDTEVANLSVNEEMKLVDSKHGQLIKNFTILQPPGPRGP KGRSQSPGTPGNGQKGEKGPPEGPAGERGTIGPVGPPGERGSKGSGQPK GSRSPGPKQPGSGDPGPPGPKDGLPGPPGPPGQGLQGTGVEGPGVPPGLPG LPVPGMPGPKPPGPPGSGAMEPLALQNEPTPASVNGCPEPHWKNFTDKYVFSLE KEILEDALKFCEKSSHLVFTNSREEQWIKKHTVGRSHWIGLTDSPQSEMKWLDG SPVDYKNWKAKQPDNMGSGHGPGECDAGLIYAGQWDPQCDDEINNFICEKREAVPSS IL"

polyA_signal 3246..3251
 /gene="srcl"

BASE COUNT 1013 a 757 c 759 g 762 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.66e-75 Length: 3291
 Score: 2330.00 Matches: 415
 Percent Similarity: 96.06% Conservative: 24
 Best Local Similarity: 90.81% Mismatches: 18
 Query Match: 92.46% Indels: 0
 Db: 10 Gaps: 0

US-09-763-712a-2_COPY_91_547 (1-457) x AB038519 (1-3291)

QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
 Db 932 ATGAATAGCCAGCTGCTCAATTCACAGGTGAGTGGCAACATTACCACTATCTCCACAG 991

QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
 Db 992 GCCACGACGACGCTGAAACACCTTCAGGACTTACACAGGATACAGAAATAGAACA 1051

QY 41 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 60
 Db 1052 GCTGTCAAGTTCCAGCAACTTCAGGAACGCTTCCAGGTCTTTGAGACAGATATTGTGAAC 1111

QY 61 IleIleSerAsnIleSerTyrThrAlaHisLysLeuArgThrLeuThrSerAsnLeuAsn 80
 Db 1112 ATCATTAGCAACATCAGCTACAGCCCATCACCCTGAGGACACTGACCACTGATCTGAAT 1171

QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
 Db 1172 GATGTTTGGACCACTGCACAGACACCTTGACACAGACACAGGATGACCTGACCTCTCTG 1231

QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
 Db 1232 AATAACACACTAGTCAACATCGCTTGGATTCTTATTTCTCTCAGGATGACGACGACATG 1291

QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
 Db 1292 ATGAGTCAAGTTAGACACTGAAGTGGCCACTTATCAGTTGGTTATGAAGAGATGAAGA 1351

QY 141 LeuValAspSerLysHisGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
 Db 1352 CTGTTGACTCCAAAGCACGCTCAGCTCATCAAGACTTTTACCATCTTACAAGGTCTCTCT 1411

QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProGlyAsn 180
 Db 1412 GCGCCACAGAGTCCAAAAGGTGACAGAGGATCTCAGGAGCACCTGCTGCTCAACTGGCAAC 1471

QY 181 LysGlyGlnLysGlyLysGlyGluProGlyGluProGlyProGlyProAlaGluArgGly 200
 Db 1472 AAGGACACAAAGGAGAGGAGGAGGAGCTGGTCCACCTGGCCCTCGGGTGTAGAGGGC 1531

QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
 Db 1532 ACAATTGGACAGTCCGCCCTCTCTGGAGAGCGTGGCGAGCAAGGATCCAAAGGCTCACAG 1591

QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
 Db 1592 GGTCCCAAAGGATCTCGTGGTCCCGAGGAAAGCCCTGGCCCTCAAGGACCTAGTGGGGAC 1651

QY 241 ProGlyProGlyProGlyLysGlyGluProGlyLysGlyProGlyProGlnGlyProGly 260
 Db 1652 CCAGGACCAAGGTCACCAAGGAGGAGTGGTCCACCTGGCCCTCAGGCGCTCTCTGGC 1711

QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
 Db 1712 TTCCAGGGACTACAGGGCACTGTGGGTGAGCCTGGAGTACCTGGACCTCGGGGTTGCCA 1771

QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
 Db 1772 GGTTCGCCAGGGTGCAGGCATGCTGGCCCTTAAGGGACCACTGGCCCTCCAGGCCCC 1831

QY 301 SerGlyAlaValValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
 Db 1832 TCAGGACCAATGAGCCATTGGCTCTGCAGAAATGAACCAACCCACGATCAGAGGTCAAC 1891

QY 321 GlyCysProProHisTrpLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
 Db 1892 GGATGTCGCCCTCACTGGGAAGAACTTCACAGATAAATGCTACTATTTTTCATTGGAAAA 1951

QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIle 360
 Db 1952 GAAATTTTGAAGATGCTAAGCTTTTCTGTGAAGACAAATCTTCCCATCTCGCTTTTTCATA 2011

QY 361 AsnThrArgGluGluGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
 Db 2012 AACTCAAGAGAAGACAGCAATGGATAAAAAGCATACCTGGGGAGAGAAAGCCATTGG 2071

QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
 Db 2072 ATCGCCCTCACAGACTCACACAGAGAAAGCGAATGGAAGTGGCTAGACGGGTCACTGTT 2131

QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGly 420
 Db 2132 GATTACAAAACCTGGAAGCTGGCAACACCATTAAGTGGGAGGAGTGGCCATGGCCAGGA 2191

QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
 Db 2192 GAAGACTGTGCTGGCTTGTATTTACGACGAGACAGTGGATGACTTCAGTGTGATGAATC 2251

```
QY 441 AsnAspHeIleCysGluLysAspArgGluThrValLeuSerSerAlaLeu 457
|||||
Db 2252 AATAACTTATTTGTGAGAGGAAGGAGGAGGAGTACCATCATCATATTA 2302
|||||
RESULT 8
HS0802985 1886 bp mRNA linear PRI 20-MAR-2002
LOCUS Homo sapiens mRNA; cDNA DKFZp547G1215 (from clone DKFZp547G1215).
DEFINITION AL713657
ACCESSION AL713657
VERSION AL713657.1 GI:19584339
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1886)
AUTHORS Koehler,K., Beyer,A., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFZp547G1215) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
FEATURES
Location/Qualifiers
1..1886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18p11.3"
/clone="DKFZp547G1215"
/tissue_type="brain"
/clone_lib="547 (synonym: hfbr1). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="fetal"
1..1886
/gene="DKFZp547G1215"
<1..1256
/gene="DKFZp547G1215"
/note="similarity to asialoglycoprotein receptor (Mus
musculus)"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAD28466.1"
/db_xref="GI:19584340"
/translation="AIKENOLEERFQLPETDIVIINISVTAHLRLTSLNLEVRT
TCTPTLTQDLSLNNTLANIRLDSVSLRMOQDLMSRLDTEVALSVIMEKLV
DSKHGLIKNTILQGPGRGDRSQPGPTGNKQKGEKGEPRGPAGERG
PIGAPPPERGKSGKSGPKSGRSGFGPGGSGDGPFGPKGELGPGQP
PGFGLQGTGEPGVPGLPGPLGPGMPKPGPPGPGPSAVPLALONETPA
PEDNGCPHKNFTDKCYFVSEKIEIFEDALFKEDSKSLVFINTRREQWIKQMV
GRESHWIGLTDSERENKWLDTGTPSPYKNNKAGQPDNWHGHGPGEDCAGLIVAGW
NDFOCEDVNNFTICEKDRVTLSSAL"
1838..1843
polyA_signal
/gene="DKFZp547G1215"
1869
polyA_site
/gene="DKFZp547G1215"
BASE COUNT 569 a 444 c 445 g 428 t
ORIGIN
Alignment Scores:
Pred. No.: 6,08e-75 Length: 1886
Score: 2309.00 Matches: 416
Percent Similarity: 99.76% Conservative: 0
Best Local Similarity: 99.76% Mismatches: 1
Query Match: 91.63% Indels: 0
DB: 9 Gaps: 0
```

```
US-09-763-712a-2_copy_91_547 (1-457) x HSM802985 (1-1886)
QY 41 AlaIleLysPheAsnGlnLeuGluArgPheGlnLeuPheGluThrAspIleValAsn 60
|||||
Db 3 GCCATCAAGTTCAACCACTGGAGGAAACGCTTCCAGCTCTTTGAGACGATATTGTGAAC 62
|||||
QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
|||||
Db 63 ATCATTAGCAATATAGTTACACAGCCACCACCTCGGACGCTGACCAAGTCTAAAT 122
|||||
QY 81 GluValArgThrCysThrAspThrLeuThrLysHisThrAspLeuThrSerLeu 100
|||||
Db 123 GAAGTCAGACCACTTGCACATACCTTACCAACACACAGATGATCTGACCTCCTTG 182
|||||
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
|||||
Db 183 AATAATACCTCGCCCAACATCGTTTGGATTCTGTTTCTCTCAGGATGCAACAAGATTG 242
|||||
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
|||||
Db 243 ATGAGTCCAGGTTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGAAGAAATGAAG 302
|||||
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAsnPheThrIleLeuGlnGlyProPro 160
|||||
Db 303 CTAGTAGACTCCCAAGCATGGTCAGCTCATCAAGAATTTTACAATACTACAAGGTCCACCG 362
|||||
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProThrGlyAsn 180
|||||
Db 363 GGCCTCCAGGGTCCCAAGAGGTACAGAGATCCAGGAGACCCCTGCTGCCCACTGGCAAC 422
|||||
QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGlyGluArgGly 200
|||||
Db 423 AAGGACACAAAGGAGAGAAAGGGGAGCCTGGACCACTTGGCCCTCGCGGTGAGAGAGGC 482
|||||
QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
|||||
Db 483 CCAATTGGACCACTGCTCCCGCGGAGAGCGTGGCGCAAGAGATCTAAAGGCTCCCAAG 542
|||||
QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlyProGlyProSerGlyAsp 240
|||||
Db 543 GGCCTCAAGAGCTCCCGTGGTCCCTGGAGAGCCCGCCCTCAGGGCTCCAGTGGGGAC 602
|||||
QY 241 ProGlyProGlyProGlyProGlyLysGlyGluGlyLeuProGlyProGlyProGlyProGly 260
|||||
Db 603 CCAGGCCCCCGGGCCACCAGCAAGAGGAGACTTCCCGGGCCCTCAGGGCCCTCTCTGGC 662
|||||
QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
|||||
Db 663 TTCCAGGGACTTTCAGGGCACCGCTTGGGGAGCCTTGGGGTGCCTGGACCTCGGGAGTGC 722
|||||
QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
|||||
Db 723 GCTTTCCTTGGGTACCAAGCATGCCAGGCCCAAGGGCCCCCGCCCTCTCTGGCCCA 782
|||||
QY 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
|||||
Db 783 TCAGGAGCGGTGCTGCTCCCTGGCCCTGCAGATGAGCAACCCGCGCACCGGAGGACAAT 842
|||||
QY 321 GlyCysProProHisThrLysAsnPheThrAspLysCysTyrTyrPheSerValGluLys 340
|||||
Db 843 GGTGCTCCCGCCCTCACTGGAAGAACTTCACAGACAAATGCTACTATTTTTCAGTTGAGAA 902
|||||
QY 341 GluIlePheGluAspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIle 360
|||||
Db 903 GAAATTTTGGAGATGCAAAAGCTTTTCTGTGAAGACAAAGTCTTCACATCTGTGTTTCATA 962
|||||
QY 361 AsnThrArgGluGlnGlnIntrPpIleLysLysGlnMetValGlyArgGluSerHisTrp 380
|||||
Db 963 AACACTAGAGAGGAACAGCAATGGATAAAAAACACAGATGGTAGGAGAGAGAGCCACTGG 1022
|||||
QY 381 IleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerPro 400
|||||
Db 1023 ATCGGCTTCACAGACTCAGAGCGTGAATGAAATGGAAGTGGCTGGATGGGACATCTCCA 1082
|||||
```



```

QY 401 AspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyProGly 420
Db 1083 GACTACAAAAATGGAAAGCTGGACAGCGGATAACTGGGGTCAATGCCATGGCCAGGA 1142
QY 421 GluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrpAsnAspPheGlnCysGluAspVal 440
Db 1143 GAAGACTGTGCTGGGTGATTTATGTCTGGGCAGTGGAAACGATTTCCAAATGTGAAGACGTC 1202
QY 441 AsnAsnPheIleCysGluLysAspArgGluThrValIleSerSerAlaLeu 457
Db 1203 AATAACTTATTTCGCAAAAAGACAGGAGACAGTACTGTCTCATCTGCAATTA 1253

RESULT 9
AB052103
LOCUS Homo sapiens SRCL mRNA for scavenger receptor with C-type lectin
DEFINITION type II, complete cds.
ACCESSION AB052103
VERSION AB052103.1 GI:13365552
SOURCE Homo sapiens CDNA to mRNA.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Nakamura,K., Funakoshi,H., Miyamoto,K., Tokunaga,F. and Nakamura,T.
TITLE Molecular cloning and functional characterization of a human
        scavenger receptor with C-type lectin (SRCL), a novel member of a
        scavenger receptor family
JOURNAL Biochem. Biophys. Res. Commun. 280 (4), 1028-1035 (2001)
MEDLINE 21092718
REFERENCE 2 (bases 1 to 4330)
AUTHORS Nakamura,K. and Nakamura,T.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-2000) Keiji Nakamura, Osaka University Graduate
        School of Medicine, Division of Biochemistry, Biomedical Research
        Center, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
        (E-mail:knakamura@onbich.med.osaka-u.ac.jp, Tel:81-6-6879-3783,
        Fax:81-6-6879-3789)
FEATURES             Location/Qualifiers
     source            1..4330
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene              1..4330
                     /gene="SRCL"
     CDS                3..1871
                     /gene="SRCL"
                     /codon_start=1
                     /product="scavenger receptor with C-type lectin type II"
                     /protein_id="BAB39148.1"
                     /db_xref="GI:13365553"
                     /translation="MKDDFAEEVEVPSFGKRFGEHGTQCPKHNNAKFSITLLY
                     ILCALLTITVAIGLYKVEKNDVNGMETSRQYDDKLITAVESDLKLIGDOTGKKAI
                     STNSLESTFRSDILDLRQRLREITEKTSKNDLLEKQASGDALVDROSQKETLENN
                     SPLITVNTKLOAYNYTNLQDDTSVLOGNQNOMYSHNVVIMNINLNTQVOQRN
                     LITNQRVDVTSQAIQRIKNDQNLQVFTQAKKDDTLKKEVQSLOTLAANNSALA
                     KANNLDLDMNSQLSFTGQMENITISQANEONLKDLQHLKDAENRTAIKFNOLKE
                     RFLPETFIVNISNYSVTAHLRTLTLNLEVRVTTDITLTKHTDDTLTLNNTLANI
                     RLDSVLRMOODLRSLDTEVANLSVIMEKMLVDSKHGQLIKNFTILQPPGPRGP
                     RGDSSQPGPTGNKQKGEKGEPPGAGERPIGPAGPPGERGKSGSGGPK
                     GSRGSPGKPGQSGPPGPGKGLPGPPGPPGFGLOGTGVGEVPGPGRGLPG
                     LPVPGMPGKPGPPGPGPSGAVPLALQNEPTPAEDNSKSPKSLQPGGGSACA"
BASE COUNT 1186 a 1065 c 1035 g 1044 t
ORIGIN
Alignment Scores:
Pred. No.: 2,41e-72 Length: 4330
Score: 2244.00 Matches: 457
Percent Similarity: 46.21% Conservative: 0
Best Local Similarity: 46.21% Mismatches: 0
Query Match: 89.05% Indels: 532
DB: 9 Gaps: 1

```

```

US-09-763-712a-2_copy_91_547 (1-457) x AB052103 (1-4330)
QY 1 MetAsnSerGlnLeuAsnSerPheThrGlyGlnMetGluAsnIleThrThrIleSerGln 20
Db 858 ATGAACAGCCAGCTCAACTCATTCACAGGTACAGTACAGAACATCACCATATCTCTCAA 917
QY 21 AlaAsnGluGlnAsnLeuLysAspLeuGlnAspLeuHisLysAspAlaGluAsnArgThr 40
Db 918 GCCAACGACGACAGAACCTGAAAGACCTGCAGGACTTACACAAAGATGCAGAGAATAGAACA 977
QY 41 AlaIleLysPheAsnGlnLeuGluGluArgPheGlnLeuPheGluThrAspIleValAsn 60
Db 978 GCCATCAAGTTCAACCAACTGGAGGAAAGCTTCCAGCTCTTTGAGACGGATATGTGAAC 1037
QY 61 IleIleSerAsnIleSerTyrThrAlaHisHisLeuArgThrLeuThrSerAsnLeuAsn 80
Db 1038 ATCATTAGCAATATCAGTTACACAGCCACCCCTCGGAGCTGACCAATCTAAAT 1097
QY 81 GluValArgThrThrCysThrAspThrLeuThrLysHisThrAspAspLeuThrSerLeu 100
Db 1098 GAAATCAGGACCACTTGCACAGATACCTTACCACAAACACACAGATGATCTGACCTCCTTG 1157
QY 101 AsnAsnThrLeuAlaAsnIleArgLeuAspSerValSerLeuArgMetGlnGlnAspLeu 120
Db 1158 AATAATACCTCGCCCAACATCCGTTTGGATTCTCTCTCAGGATGCAACAAAGATTG 1217
QY 121 MetArgSerArgLeuAspThrGluValAlaAsnLeuSerValIleMetGluGluMetLys 140
Db 1218 ATGAGGTCGAGGTAGACACTGAAGTAGCCAACTTATCAGTGATTATGGNAGAAATGAAG 1277
QY 141 LeuValAspSerLysHisGlyGlnLeuIleLysAspPheThrIleLeuGlnGlyProPro 160
Db 1278 CTAGTAGACTCCAAAGCATGGTCAGCTCATCAAGATTTTACAATACTACAAGGTCCACCG 1337
QY 161 GlyProArgGlyProArgGlyAspArgGlySerGlnGlyProGlyProGlyProGlyAsp 180
Db 1338 GGCCCCAGGGGTCCAGAGGTGACAGAGGATCCAGGGACCCCTCGCCCACTGGCAAC 1397
QY 181 LysGlyGlnLysGlyGluLysGlyGluProGlyProGlyProGlyProAlaGlyGluArgGly 200
Db 1398 AAGGGACAGAAAGGAGAGAGGGGAGCCTGGACCACTGGCCCTCGGGGTGAGAGAGGC 1457
QY 201 ProIleGlyProAlaGlyProGlyGluArgGlyGlyLysGlySerLysGlySerGln 220
Db 1458 CCAATTGGACCACTGGTCCCCCGGAGAGCGTGGCGGCAAAAGGATCTAAAGGCTCCAG 1517
QY 221 GlyProLysGlySerArgGlySerProGlyLysProGlyProGlnGlyProSerGlyAsp 240
Db 1518 GGCCCCAAAGGCTCCCGTGGTTCCCTTGGGAAGCCCGGCCCTCAGGGCCCCCAGTGGGAC 1577
QY 241 ProGlyProGlyProGlyProGlyLysGlyGluGlyLeuProGlyProGlnGlyProGly 260
Db 1578 CCAGGGCCCCCGGGGCCACACAGGCAAGAGGAGTCCCGGGCCCTCAGGGCCCTCTCTGC 1637
QY 261 PheGlnGlyLeuGlnGlyThrValGlyGluProGlyValProGlyProArgGlyLeuPro 280
Db 1638 TTCCAGGGACTTCAGGGCACCGTGTGGGAGCGCTGGGTGCCTGGACCTCGGGGACTGCCA 1697
QY 281 GlyLeuProGlyValProGlyMetProGlyProLysGlyProGlyProGlyProGlyPro 300
Db 1698 GGCTTGCTCTGGGTACCAAGGATGCCAGGCCCCCAAGGGCCCCCGGCCCTCTCTGCCCA 1757
QY 301 SerGlyAlaValProLeuAlaLeuGlnAsnGluProThrProAlaProGluAspAsn 320
Db 1758 TCAGGAGCGGTGGTGGCCCTCGCCCTGCAAGATGACCAACCCCGGCACGGAGACAAT 1817
QY 320 ----- 320
Db 1818 AGTAAGTCCAAGCCCTCCCTCCAGCCAGGGGACAGGGCAGTGCATGTGCTTGAGACTGA 1877
QY 320 ----- 320

```


Db 1878 GCCAACCCCACTGTGTGGCCCCCAACCATGTATTTTGTGTTTGTGTTGAGAC 1937
QY 320 ----- 320
Db 1938 AGGGTCTCCTCTGTTGCCAGGCTGGAGTGCAGTGGCGTGATCTCGGCTCACTGCAACC 1997
QY 320 ----- 320
Db 1998 CCCACATCCGGGTTCAAGTGATTCTCCTGCTTAGCCTCCAGTAGCTGGGAATACAA 2057
QY 320 ----- 320
Db 2058 GCGTGTGCCCCACAGCCAGCTAATTTTGTATTTTGTATTTAGTAGAGAGGGGTTTCACAGTGT 2117
QY 320 ----- 320
Db 2118 TGGCAGGGTGTTCACAACTTCCTGGCCTCAAGTCACCTACCTAGTTTCCCAAGT 2177
QY 320 ----- 320
Db 2178 GCTGGGAATACAGGTGTGAACACACACCGTTTTTGTCCCGGCTGCATATTTATTAACAGC 2237
QY 320 ----- 320
Db 2238 AAAGCATCATAGGGGTTGGCACAGGAGCCCTGCTGCAGAACAGGCCCTCCATACCAAGC 2297
QY 320 ----- 320
Db 2298 ATGTGCTCTCTCAACGCCATTTGGGTTTTTGTAGTGTCTATTTTGTATTTATTTATGT 2357
QY 320 ----- 320
Db 2358 GTGTATGTATGGGACAGGGTCTCACTCTGTCAACCAGGCTGGAATGACGAAGTGAAGTAG 2417
QY 320 ----- 320
Db 2418 TGCAACATAGCTCATTCACAGCTTTGACCTCTGCGGCTTAGTGATCTTCCCATCAGAC 2477
QY 320 ----- 320
Db 2478 TCCTCGTAGTGGACACAGGGTATGCCACCACACCTGGCTAAATTTTTTATTTTCT 2537
QY 320 ----- 320
Db 2538 TTTTGTAGAAACAGGGTCTTGCTATGTTCGCCAGGCTGGTCTTGAACCTCTGGCCTCAAG 2597
QY 320 ----- 320
Db 2598 CAACCTCCTGCTCAGTCTTGCAAGTGCCTGGGATTCAGGTGGGAGCCACTGCACCTG 2657
QY 320 ----- 320
Db 2658 GCTCTTGCACTTGTTATTTATAGGAGCCAGCCCTCAGAAGACAAACACATGCAAAACA 2717
QY 320 ----- 320
Db 2718 TTTCTTCAGGAAGCCTGCTGGCGCCCTGAGCGCTCTCCCTCGTTTTACATCTCCAC 2777
QY 320 ----- 320
Db 2778 TCTGGAGATAGTGGCTCCCTCGGTAGTCCATCAACATGTGGATGGAACCCCTCTGCT 2837
QY 320 ----- 320
Db 2838 ATTGATGACTACATTTAATTTCTTTTGTAGAAAGTGTTAGATGTTACGGTGCCTTCTAG 2897
QY 320 ----- 320
Db 2898 TCTTTTGATAATAAACCTTTGCAAAACTCGAATGTTTGTGTAGCAACCAAGCCTATTT 2957
QY 320 ----- 320
Db 2958 TTCCTCCTTTGACCAACCCCTATTTACTTGTATTTTGGTGGTGGAGCCCGGGTTATT 3017

QY 320 ----- 320
Db 3018 TTTGTTGATCCGTGAGCCCATTTTGTGGTGTGGGTTCTCACATCCAAGTGTCTCTGGGTT 3077
QY 320 ----- 320
Db 3078 CCGTATGATTCAGACTGACGCTGAACATCTGTGCTGCCCTGCTTCTTGTGTTGCTTC 3137
QY 320 ----- 320
Db 3138 TGTCAATTTCTGTCAPTGTGAAGACGGATCCAACACAGAAGAACAAAGACCCAGTCCAAG 3197
QY 320 ----- 320
Db 3198 ATCCATCCAGATTATGTGGTTTCGGCTTGTGCTCTGCAGTCCACTTTGGGTCACTGTGC 3257
QY 320 ----- 320
Db 3258 TCGAGACAGTCCGTTGAACGTGAGCTGTGGGCTCGTGCAGGAGCGCTCAGGCTGCTCCC 3317
QY 320 ----- 320
Db 3318 AGCGGACCAAGTATTGGCCCTCTGAACACAGCTGAAGAAAGGACAGAGTGTCTTCTGTG 3377
QY 321 -----
Db 3378 ATCTTTTAATCAATGTTTTCACACTTTTGCATTCCTAGGCTGCCGCCCTCACATGGGAAGAC 3437
QY 329 PheThrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAspAlaLysLeu 348
Db 3438 TTCACAGACAAATGCTACTATTTTTCAGTTGAGAAGAAATTTTTCAGGATGCAAGCTT 3497
QY 349 PheCysGluAspLysSerHisLeuValPheIleAsnThrArgGluGluGlnGlnTrp 368
Db 3498 TTCGTGAAGACAAGTCTTCACATCTTGTTCATAAACACTAGAGAGAACAGCAATGG 3557
QY 369 IleLysLysGlnMetValGlyArgGluSerHisThrIleGlyLeuThrAspSerGluArg 388
Db 3558 ATAAAAAACAGATGGTAGGAGAGAGAGCCACTGATCGGCCCTCACAGACTCAGAGCGT 3617
QY 389 GluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTrpLysAlaGly 408
Db 3618 GAAATGAATGGAAGTGGCTGGATGGGACATCTCCAGACTACAAAATTTGGAAGCTGGA 3677
QY 409 GlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAspCysAlaGlyLeuIleTyr 428
Db 3678 CAGCCGGATAACTGGGGTCATGGCCATGGCCAGGAGAGACTGTGCTGGGTGATTAT 3737
QY 429 AlaGlyGlnTrpAsnAspPheGlnCysGluAspValAsnAsnPheIleCysGluLysAsp 448
Db 3738 GCTGGCAGTGGAAAGCATTTCCAATGTGAAGAGCTCAATAACTTCATTTGCAAAAGAC 3797
QY 449 ArgGluThrValLeuSerSerAlaLeu 457
Db 3798 AGGGAGACAGTACTGTCTATCTGCATTA 3824

RESULT 10

AC112416

LOCUS

DEFINITION

AC112416

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC112416 130763 bp DNA linear HTG 13-JUL-2002
Rattus norvegicus clone CH230-350K4, *** SEQUENCING IN PROGRESS
***, 28 unordered pieces.

AC112416 GI:21743403

HTG; HTGS_PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 130763)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Aisbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Eathwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, K., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louiseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 130763)
Worley, K.C.

Direct Submission
Submitted (21-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 130763)
Worley, K.C.

Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:20303233.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRRG
Center clone name: CH230-350K4
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 113686 bases at least Q40
Consensus quality: 115729 bases at least Q30
Consensus quality: 117044 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1103: contig of 1103 bp in length
* 1104 1203: gap of unknown length
* 1204 3068: contig of 1865 bp in length
* 3069 3168: gap of unknown length
* 3169 4755: contig of 1587 bp in length
* 4756 4855: gap of unknown length
* 4856 7046: contig of 2191 bp in length
* 7047 7146: gap of unknown length
* 7147 8651: contig of 1505 bp in length
* 8652 8751: gap of unknown length
* 8752 10001: contig of 1250 bp in length
* 10002 10101: gap of unknown length
* 10102 11278: contig of 1177 bp in length
* 11279 11378: gap of unknown length
* 11379 13651: contig of 2273 bp in length
* 13652 13751: gap of unknown length
* 13752 15942: contig of 2191 bp in length
* 15943 16042: gap of unknown length
* 16043 17619: contig of 1577 bp in length
* 17620 17719: gap of unknown length
* 17720 20301: contig of 2582 bp in length
* 20302 20401: gap of unknown length
* 20402 23598: contig of 3197 bp in length
* 23599 23698: gap of unknown length
* 23699 26125: contig of 2427 bp in length
* 26126 26225: gap of unknown length
* 26226 30263: contig of 4038 bp in length
* 30264 32964: gap of unknown length
* 32965 33064: contig of 2601 bp in length
* 33065 36455: gap of unknown length
* 36456 40431: contig of 3876 bp in length
* 40432 40531: gap of unknown length
* 40532 42635: contig of 2104 bp in length
* 42636 42735: gap of unknown length
* 42736 46230: contig of 3495 bp in length
* 46231 46330: gap of unknown length
* 46331 49593: contig of 3263 bp in length
* 49594 49694: gap of unknown length
* 49694 55037: contig of 5343 bp in length
* 55037 55137: gap of unknown length
* 55137 60936: contig of 5800 bp in length
* 60937 61036: gap of unknown length
* 61037 69030: contig of 7994 bp in length
* 69031 69130: gap of unknown length
* 69131 79184: contig of 10054 bp in length
* 79185 79284: gap of unknown length
* 79285 91807: contig of 12523 bp in length
* 91808 91907: gap of unknown length
* 91908 106422: contig of 14515 bp in length
* 106423 106522: gap of unknown length
* 106523 119310: contig of 12788 bp in length
* 119311 119411: gap of unknown length
* 119411 130763: contig of 11353 bp in length.

FEATURES

Location/Qualifiers
1..130763
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-350K4"

BASE COUNT 38110 a 25956 c 25888 g 38088 t 2721 others
ORIGIN

Alignment Scores:

Pred. No.: 1.39e-27 Length: 130763
Score: 1021.50 Matches: 215
Percent Similarity: 41.4% Conservative: 16

| | | | |
|------------------------|--------|-------------|-----|
| Best Local Similarity: | 38.60% | Mismatches: | 37 |
| Query Match: | 40.54% | Indels: | 289 |
| DB: | 2 | Gaps: | 5 |

US-09-763-712A-2_COPY_91_547 (1-457) x AC112416 (1-130763)

| | | | |
|----|-------|---|-------|
| QY | 153 | PhethrleleuGlnGlyProProGlyProArgGlyProArgGlyAspArgGlySerGln | 172 |
| Db | 13787 | TTTTCTTTATGTTTATGTTCTCTCTGGCCCCAGAGGTCCTCAAAAGGTGACAGAGGATCTCAG | 13846 |
| QY | 173 | GlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGlyGluProGlyPro | 192 |
| Db | 13847 | GGACCGCTGGTCCAACCTGCCAACAGGGACAGAAAGGAGAGAGGAGACCTTGGTCCA | 13906 |
| QY | 193 | ProGlyProAlaGlyGluArgGlyProIleGlyProAlaGlyProProGlyGluArgGly | 212 |
| Db | 13907 | CCTGGCCCTGGGTGAGCGAGGCACAAATTGGACCACTAGGCCCCCTCTGGAGAGCGGTGGC | 13966 |
| QY | 213 | GlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySerProGlyLysPro | 232 |
| Db | 13967 | AGCAAAAGATCCAAGAGCTCACAGGGTCCCAAGAGATCACTGGGTCCCCAGGAAGCCT | 14026 |
| QY | 233 | GlyProGlnGlyProSerGlyAspProGlyProProGlyProProGlyLysGluGlyLeu | 252 |
| Db | 14027 | GGCCCTCAAGGGCCTAGTGGAGACCTTGGACCAACCAAGTCCACCAAGGATGGACTC | 14086 |
| QY | 253 | ProGlyProGlnGlyProProGlyPheGlnGlyLeuGlnGlyThrValGlyGluProGly | 272 |
| Db | 14087 | CCTGGCCCTCAGGGCCCTCTCTGGCTTCCAGGGACTACAGGGCACTGTGGTGAACCTGGA | 14146 |
| QY | 273 | ValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMetProGlyProLys | 292 |
| Db | 14147 | GTACCTGGACCTCGGGGGCTACCAGGCTTACCAGGGTGCACGCACTGCCTGGCCCCAAG | 14206 |
| QY | 293 | GlyProGlyProProGlyProSerGlyAlaValProLeuAlaLeuGlnAsnGlu | 312 |
| Db | 14207 | GGACCACTGGCCCCCAGGTCCCTCAGGAGCAATGGAGCGTGGGTCTGCAGAAATGAA | 14266 |
| QY | 313 | ProThrProAlaProGluAspAsp- | 320 |
| Db | 14267 | CCAAACCCAGCATCAGAGGTCAACGGTAAGTCCAAATCATCTTTTTCAGCCAGAAATAAG | 14326 |
| QY | 320 | ----- | 320 |
| Db | 14327 | GCACTGTGTAGGCTTGGCTAAACACAGACCTCAGGAGTCCCTAATTACACTAATATCAG | 14386 |
| QY | 320 | ----- | 320 |
| Db | 14387 | GGACCCCTATGTGTTCTCTGGATGCCCTGGGTGCTGTCACTCAGGACCCGTGTGTTCTTGCT | 14446 |
| QY | 320 | ----- | 320 |
| Db | 14447 | TAAACATTATCAGGACTCTACTCATTTAGATGACCGAGCGCCCTCAGAAAGATACCTATACC | 14506 |
| QY | 320 | ----- | 320 |
| Db | 14507 | CCAAGTACTTATTCAAGTATCTCGACGGCTGCTCCCTGACCACTCTTCTTCTGTCTT | 14566 |
| QY | 320 | ----- | 320 |
| Db | 14567 | TCTTCTTTCTCTAGAAACAGACCAATTCATCAATTAATCCATCAGTATGAGATGGGG | 14626 |
| QY | 320 | ----- | 320 |
| Db | 14627 | TTTCTGTGATAGACCACTCTGGTTTTTGGCTTAAAGAGCTGCTGTGGTGTCTGCTTTTT | 14686 |
| QY | 320 | ----- | 320 |
| Db | 14687 | GACCATTAAGTCTTCACAAAACCTCTGTCTAGTTAGGNAACAAGCCATGTGGTCACTCA | 14746 |
| QY | 320 | ----- | 320 |
| Db | 14747 | TGCTGGTCCATGAACATTTGGTTCTGAAGTACACTTGTTTACAAGCTCAGCTGATCCAAG | 14806 |

| QY | 320 | ----- | 320 |
|------------|-------|--|------------------------------------|
| Db | 14807 | ACTGCAATGAGTATCAGGGCCCCACCCACTTCTCTTCTGTCTATATAAAGCCTAAC | 14866 |
| QY | 320 | ----- | 320 |
| Db | 14867 | CCAAACCCCAATCACACGCCCTTCCACACTGTGTGGTGTTCACATTTGTGTACTGTAAATCC | 14926 |
| QY | 320 | ----- | 320 |
| Db | 14927 | ACAACTGTGAACATCTGTCCTATGTTCTGTACAGGACACATCCATTGGCATGCAAGCAGTG | 14986 |
| QY | 320 | ----- | 320 |
| Db | 14987 | TGAGCTGTCACTTTTGGTATCTTAAACATAGCTGAGAAATGCAGAGTGCCTTTTTTTTT | 15046 |
| QY | 321 | -----GlyCysProHisTrpLysAsnPheThr | 330 |
| Db | 15047 | AGATCAACATTTAAAAACGTTACATTCCTAGGATGTCGCCCTCACTGGAAGAACTTCACA | 15106 |
| QY | 331 | AspLysCysTyrTrpPheSerValGluLysGluLlePheGluAspAlaLysLeuPheCys | 350 |
| Db | 15107 | GATAAATGTTACTATTTTCGGTGAGAAAGAAATTTTGAAGATGCTAAAGCTTTTCTGT | 15166 |
| QY | 351 | GluAspLysSerHisLeuValPheIleAsnThrArgGluGluGln | 366 |
| Db | 15167 | GAAGACAATCTCCCATCTCGTTTCATAAATCAAGAGAGAAAGATGATGCAGCTG | 15226 |
| QY | 367 | ---GlnTrpIleLysLysGlnMetValGlyArg | 382 |
| Db | 15227 | CATACCTGGTTGGAGAAAACATATAATTAGAGATAAGTTATACTTGAATAATGCATGCCT | 15286 |
| QY | 383 | LeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyr | 402 |
| Db | 15287 | TTGAGGGAGTCAAGAGATCAAAAATTCATGATCTCATAAACAAAGAAAGCTGCAGTTG | 15346 |
| QY | 403 | LysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGluAsp | 422 |
| Db | 15347 | -----CATGGCTTTATCCAGGCATCGCA | 15370 |
| QY | 423 | CysAlaGlyLeu | 432 |
| Db | 15371 | ATATGTGAATGTCTGCTTCTTAGGCCCCACAGATATATAGTCCACCTGG | 15421 |
| RESULT | 11 | | |
| AC114677 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| AC114677 | | 182029 bp | DNA linear |
| HTG | | | HTG 10-JUN-2002 |
| KEYWORDS | | | WORKING DRAFT SEQUENCE, 10 ordered |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| REFERENCE | | | |
| AUTHORS | | | |

| | | | | | |
|------------|---|--|----------------------------|--------|-----------------|
| QY | 325 | HistTrpLysAsnPhetThrAspLysCysTyrTyrPheSerValGluLysGluIlePheGlu | 344 | | |
| Db | 47546 | CAC TGG AAG AACT TCA CAG AATA AAT GC TACT ATAT TTTT CATT GGA AAG AAT TTTT TCG AA | 47605 | | |
| QY | 345 | AspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArgGlu | 364 | | |
| Db | 47606 | GAT CTA AGC TTTT CTG TGA GACA AAT CTCC CAT CTC GTT TTT CAT AATA AACT CAAG AGAA | 47665 | | |
| QY | 365 | GlucInGln-----TrrpleLysLysGlnMetValGlyArgGlu | 377 | | |
| Db | 47666 | GAACAGGT-ATGATATGTCAGCTGCACACTTGGTTGGAGAAACATATAATTTAGAGATA-47723 | | | |
| QY | 378 | SerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTyrTrpLeuAspGly | 397 | | |
| Db | 47724 | AGTTATATCTTGGAAATTTGCATCCCTTTTCAGAGAGCTCAAGAGAGATCAAA-----47771 | | | |
| QY | 398 | ThrSerProAspTyrLysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHis | 417 | | |
| Db | 47772 | -----AATCATGCTCTCAATAACAA47792 | | | |
| QY | 418 | GlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrp-----AsnAspPheGln | 436 | | |
| Db | 47793 | AGGAAGGCTGCAGTTTGCATGGCTTTTCCAGGTATGAGAGATATGAATGAATTCAAA47852 | | | |
| RESULT 12 | AC102618 | 193208 bp | DNA | linear | HTG 21-AUG-2002 |
| LOCUS | AC102618 | Mus musculus clone RP23-426G16, | WORKING DRAFT SEQUENCE, 11 | | |
| DEFINITION | AC102618 | unordered pieces. | | | |
| ACCESSION | AC102618 | 2 | GI-22381604 | | |
| VERSION | HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP. | | | | |
| KEYWORDS | house mouse. | | | | |
| SOURCE | Mus musculus | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | | |
| REFERENCE | Birren, B., Nussbaum, C. and Lander, E. | | | | |
| AUTHORS | 1 (bases 1 to 193208) | | | | |
| TITLE | Mus musculus, clone RP23-426G16 | | | | |
| JOURNAL | Unpublished | | | | |
| REFERENCE | 2 (bases 1 to 193208) | | | | |
| AUTHORS | Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heard, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA | | | | |
| REFERENCE | 3 (bases 1 to 193208) | | | | |
| AUTHORS | Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Gird, S., Graham, L., Grand-Pierre, N., Hagos, B., Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M. | | | | |


```
QY 318 ----- 318
Db 98480 GAGGAATAAGCAGTGTGTAGGCTTGGCTAAACCAACCCAGAGTCCCTAATTACAC 98539
QY 318 ----- 318
Db 98540 AAAACAGCAGGACCCCTGTGTATCCCTCGGATGCTTACAGTCTGCTACAGACCCCTG 98599
QY 318 ----- 318
Db 98600 TGTTCCTGCTTAGACATATATCAGGCTCTACTCATTTAGATGAGCCCTTCAGAAGA 98659
QY 318 ----- 318
Db 98660 CCCCTGTACCCCAATACTTCTTCAAGATATTTACTACAGTGGCTGCTCCCTACCATCTCT 98719
QY 318 ----- 318
Db 98720 CCTTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 98779
QY 318 ----- 318
Db 98780 TCCAGATAGGCTTCTGTGACAGATCCACATCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 98839
QY 318 ----- 318
Db 98840 TGCCTTCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 98899
QY 318 ----- 318
Db 98900 GTGTGGTCCCTTCATGCTGTGCTTCCACAGCCAGACTTTGGTTCTGAAGTGTGCTTATTAC 98959
QY 318 ----- 318
Db 98960 AACCTCCAGCCGATCCAGCCGACAGTGTATCAGAACCCACCCACTTCTTCTTCTTCTTCTGT 99019
QY 318 ----- 318
Db 99020 CGTCATAAGCCTAGCCCAACCCAGCTCCACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 99079
QY 318 ----- 318
Db 99080 TGTGCTGTACAGTCCACATCTGAGAACATCTGCCCGGTTCTTACACAGGACACCCATTG 99139
QY 318 ----- 318
Db 99140 GCATCAAGTGTACAGGCTGTGACTTGTATCTTAAACACAGCTGAGGAAATGCAGG 99199
QY 319 ----- AspAsnGlyCysProPro 324
Db 99200 GTTCCTTCTGTGTTTTTAAATCAACAGTTAAACCTTTACATTCCTA-GGATGTCGCGCT 99258
QY 325 HisTrpLysAsnPherThrAspLysCysTyrTyrPheSerValGluLysGluPheGlu 344
Db 99259 CACTGGAAGAACTTCCACAGATAAATGCTACTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 99318
QY 345 AspAlaLysLeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArgGlu 364
Db 99319 GATGCTAGCTTTCTGTGAGACAAATCTTCCCACTCTGTTTCTTCAATCAATCAAGAGAA 99378
QY 365 GluGlnGln-----TrpIleLysLysGlnMetValGlyArgGlu 377
Db 99379 GAACAGGT-ATGATATGTCAGTGCACACTTGGTTGGAGAAACTAATATTAGAGATA- 99436
QY 378 SerHisTrpIleGlyLeuThrAspSerGluArgGluAsnGluTrpLysTrpLeuAspGly 397
Db 99437 AGTTATTCTGAAATTCATGCTCCCTTTGAGAGAGTCAAGAGATCAAA----- 99484
QY 398 ThrSerProAspTyr-LysAsnTrpLysAlaGlyGlnProAspAsnTrpGlyHisGlyHis 417
Db 99485 -----AATTCATGCTGCTCATAACAA 99505
```

```
QY 418 GlyProGlyGluAspCysAlaGlyLeuIleTyrAlaGlyGlnTrp---AsnAspPheGln 436
Db 99506 AGGAAGGCTGCAGTTGTCATGGCTTTTCCAGGTATGAGATATATGATGAATTCATAA 99565

RESULT 13
AC016128/c
LOCUS
DEFINITION
AC016128 Homo sapiens chromosome 18 clone RP11-324G2 map 18, linear HTG 09-SEP-2000
SEQUENCE, 8 unordered pieces.
AC016128
AC016128.4 GI:10046526
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 169088)
Homo sapiens chromosome 18, clone RP11-324G2
Unpublished
2 (bases 1 to 169088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguski,M., Boughey,L., Boughey,L.,
Brown,A., Castle,A., Collins,S., Collins,S., Collins,S., Collins,S.,
Cooke,P., DeArnell,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hags,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatsis,A., Klein,J.,
Lehoczky,J., Liu,C., Locke,K., Macdonald,P., Marquis,N.,
McSwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meidirim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 9, 2000 this sequence version replaced gi:6649269.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L999
Center clone name: 324.G.2
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 16522 bases at least Q40
Consensus quality: 167248 bases at least Q30
Consensus quality: 167996 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 168388; sum-of-ctg
Quality coverage: 6.2 in Q20 bases; agarose-fp
Quality coverage: 6.5 in Q20 bases; sum-of-ctgs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 17943: contig of 17943 bp in length
* 17944 18043: gap of 100 bp
* 18044 22424: contig of 4381 bp in length
```

```

* 22425 22524: gap of 100 bp
* 22525 38094: contig of 15570 bp in length
* 38095 38194: gap of 100 bp
* 38195 49220: contig of 11026 bp in length
* 49221 49320: gap of 100 bp
* 49321 66394: contig of 17074 bp in length
* 66395 66494: gap of 100 bp
* 66495 91692: contig of 25198 bp in length
* 91693 91792: gap of 100 bp
* 91793 120869: contig of 29077 bp in length
* 120870 120969: gap of 100 bp
* 120970 169088: contig of 48119 bp in length.

```

FEATURES

source

```

1. .169088
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"

```

```

/clone="RP11-324G2"
/cclone_lib="RPC1-11 Human Male BAC"

```

misc_feature

```

1. .17943
/note="assembly_fragment"
clone_end: SP6
vector_side: left"

```

misc_feature

```

18044. .22424
/note="assembly_fragment"

```

misc_feature

```

22525. .38094
/note="assembly_fragment"

```

misc_feature

```

38195. .49220
/note="assembly_fragment"

```

misc_feature

```

49321. .66394
/note="assembly_fragment"

```

misc_feature

```

66495. .91692
/note="assembly_fragment"

```

misc_feature

```

91793. .120869
/note="assembly_fragment"

```

misc_feature

```

120970. .169088
/note="assembly_fragment"

```

```

clone_end: T7
vector_side: right"

```

```

BASE COUNT 48103 a 33537 c 34088 g 52660 t 700 others
ORIGIN

```

Alignment Scores:

```

Pred. No.: 4.16e-25 Length: 169088
Score: 954.50 Matches: 191
Percent Similarity: 58.40% Conservative: 21
Best Local Similarity: 52.62% Mismatches: 53
Query Match: 37.88% Indels: 98
DB: 2 Gaps: 8

```

```

US-09-763-712a-2_COPY_91_547 (1-457) x AC016128 (1-169088)

```

```

Qy 131 AnLeuSerValIleMetGluCiuMetLysLeuValAspSerLysHisGly-----Gln 148
||||| :|::: ||||| :|:::
Db 153903 AATTTGGAGATTTTAAAGAGAAAAATTTTCATTAATTCACAGTGGATTTCAT 153844

Qy 149 LeuLeuLysAsnPhetHrLeuGlnGlyProGlyProArgGlyProArgGlyAsp 168
||||| :|::: ||||| :|:::
Db 153843 GTTGACACCTTTTCTTTATGCTTTAGTGTCCACCGGGCCCGAGAGGTGAC 153784

Qy 169 ArgGlySerGlnGlyProGlyProThrGlyAsnLysGlyGlnLysGlyGluLysGly 188
||||| :|::: ||||| :|:::
Db 153783 AGAGATCCAGGACCCCTGGCCCACTGGCACAGAGGACAGAAAGAGAGAGGGG 153724

Qy 189 GluProGlyProGlyProAlaGlyGluArgGlyProLleGlyProAlaGlyProPro 208
||||| :|::: ||||| :|:::
Db 153723 GAGCCTGGACCACTGGCCCTGGCGGTGAGAGAGGCCCAATTTGGACAGCTGGTCC 153664

Qy 209 GlyGluArgGlyGlyLysGlySerLysGlySerGlnGlyProLysGlySerArgGlySer 228
||||| :|::: ||||| :|:::
Db 153663 GGAGAGCGTGGCGGCAAGAGTCTAAAGGCTCCCGAGGCCCCCAAGGCTCCCGTGGTTCC 153604

```

```

Qy 229 ProGlyLysProGlyProGlnGlyProSerGlyAspProGlyProGlyProGlyProGly 248
Db 153603 CCTGGGAAGCCCGGCTCCAGGGCTCCAGTGGGAGCCAGGCCCCGGGGCCACAGG 153544

Qy 249 LysGluGlyLeuProGlyProGlnGlyProGlyPheGlnGlyLeuGlnGlyThrVal 268
Db 153543 AAAGAGGACTCCCGGGCCCTCAGGGCCCTCTGGCTTCCAGGGACTTTCAGGGCACCCTT 153484

Qy 269 GlyGluProGlyValProGlyProArgGlyLeuProGlyLeuProGlyValProGlyMet 288
Db 153483 GGGAGCCTGGGTGCTGGACCTCGGGGACTGCCAGGCTTGCCTGGGGTACCAGGCATG 153424

Qy 289 ProGlyProLysGlyProGlyProGlyProGlyProSerGlyAlaValProLeuAla 308
Db 153423 CCAGGCCCAAGGGCCCCCGGGCCCTCTGGCCCATCAGGAGCGGTGGTCCCTGGGCC 153364

Qy 309 LeuGlnAsnGluProThrProAlaProGluAsnGly----- 321
Db 153363 CTGCAGATGAGCCCAACCCAGCACCGGAGGACATGTAAAGTCCAAAGCCCTCCCTCCAG 153304

Qy 322 -----CysPro-ProHi 325
Db 153303 CCAGGGGACAGGCGCATGTGCTTGAGACTGAGCCCAACCCCGAGTGTGGCCCCA 153244

Qy 325 sTrpLysAsnPhetHrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAs 345
Db 153243 C-----CCATGTTATTT----- 153231

Qy 345 pAlaLysLeuPheCysGluAspLysSerSerHisLeuValPheIleAsnThrArgGluGl 365
Db 153230 -TGTTTTTGTGTTTGTGTTTGTGAGACAGGCTCTCAGTCTCTGTT----- 153192

Qy 365 uGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAs 385
Db 153192 ----- 153192

Qy 385 pSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTyrLysAsnTr 405
Db 153191 -GCCAGGCTGGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 153133

Qy 405 pLys-----AlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyGly 421
Db 153132 CAAGTGATTCTCTGCGCTTAGCCTCCCGAGTGTGGGACTACAAAGCGTGTGCCCCCACA 153073

Qy 421 uAspCysAlaGlyLeuIleTyrAlaGlyGln-----TrpAs 433
Db 153072 GCCAGCTAAATTTTGTATTTTAGTAGAGAGGGGGTTTCACCGTGTGGCAGGCTGTC 153013

Qy 433 n-----AspPheGl 436
Db 153012 TCAAACTCTGGCCTCAAGTGCACCTACCTAGTCTCCCAAGTGTGGGATTACAG 152953

Qy 436 nCysGlu 438
Db 152952 GTGTGAG 152946

RESULT 14
AP000915/c
LOCUS AP000915
DEFINITION Homo sapiens genomic DNA, chromosome 18p clone:RP11-720L2, complete
sequences.
ACCESSION AP000915
VERSION AP000915.5 GI:20334314
KEYWORDS HNG.
SOURCE Homo sapiens DNA, clone:RP11-720L2.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (1999)

```


| | | | |
|------------|--|--|-------|
| QY | 325 | strPlysAsnPhetrAspLysCysTyrTyrPheSerValGluLysGluIlePheGluAs | 345 |
| Db | 44364 | C-----CCATGTTATTT----- | 44352 |
| QY | 345 | pAlaLysLeuPheCysGluAspLysSerHisLeuValPheIleAsnThrArgGluG1 | 365 |
| Db | 44351 | -TGTGTTTGTGTTTGTGTTTGTGACAGGCTCTCACTCTGTT----- | 44313 |
| QY | 365 | uGlnGlnTrpIleLysLysGlnMetValGlyArgGluSerHisTrpIleGlyLeuThrAs | 385 |
| Db | 44313 | ----- | 44313 |
| QY | 385 | pSerGluArgGluAsnGluTrpLysTrpLeuAspGlyThrSerProAspTrpLysAsnTr | 405 |
| Db | 44312 | -GCCAGCGTGGAGTGCAGTGGCGGTGATCTCGGCTCACTGCAACCTCCATCCCGGGTT | 44254 |
| QY | 405 | pLys-----AlaGlyGlnProAspAsnTrpGlyHisGlyHisGlyProGlyG1 | 421 |
| Db | 44253 | CAAGTGATTCCTCGCTTACGCTCCAGGTAGCTGGGACTACAAGCGTGTGCCCCACA | 44194 |
| QY | 421 | uAspCysAlaGlyLeuIleLysAlaGlyGln----- | 433 |
| Db | 44193 | GCCAGCTAATTTTGTGTTATTTTAGTAGAGAGGGGTTTCACCGTGTGGCAGCTGGTC | 44134 |
| QY | 433 | n----- | 436 |
| Db | 44133 | TCAAACCTCTGGCGCTCAAGTGATCCACCTAGCTTCTCCAAAGTGCTGGGATTACAG | 44074 |
| QY | 436 | nCysGlu | 438 |
| Db | 44073 | GTGTGG | 44067 |
| RESULT | 15 | | |
| ACCESSION | AC024368 | | |
| LOCUS | Homo sapiens chromosome 11 clone RP11-179K3 map 11, LOW-PASS | | |
| DEFINITION | SEQUENCE SAMPLING. | | |
| ACCESSION | AC024368.1 | GI:7108157 | |
| VERSION | HTG; HTGS_PHASE0. | | |
| KEYWORDS | Homo sapiens | | |
| SOURCE | Homo sapiens | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| REFERENCE | 1 (bases 1 to 71044) | | |
| AUTHORS | Blirren, B., Linton, L., Nusbaum, C. and Lander, E. | | |
| TITLE | Homo sapiens chromosome 11, clone RP11-179K3 | | |
| JOURNAL | Unpublished | | |
| REFERENCE | 2 (bases 1 to 71044) | | |
| AUTHORS | Blirren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bada, F., Boguslavsky, L., Bouknight, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Largoocke, K., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihoval, T., Miranda, C., Mlangwa, V., Morrow, J., Naylor, J., Norman, C. H., O'Connor, J., O'Donnell, P., O'Neill, D., Oliver, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigliio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome | | |

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6145

Center clone name: 179_K_3

* NOTE: This record contains 90 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 699 798: 698: contig of 698 bp in length
 799 798: gap of 100 bp
 1495 1495: contig of 697 bp in length
 1496 1595: gap of 100 bp
 1596 2284: contig of 689 bp in length
 2285 2384: gap of 100 bp
 2385 3056: contig of 672 bp in length
 3057 3156: gap of 100 bp
 3157 3851: contig of 695 bp in length
 3852 3951: gap of 100 bp
 3952 4627: contig of 676 bp in length
 4628 4727: gap of 100 bp
 4728 5409: contig of 682 bp in length
 5410 5509: gap of 100 bp
 5510 6183: contig of 674 bp in length
 6184 6283: gap of 100 bp
 6284 6984: contig of 701 bp in length
 6985 7084: gap of 100 bp
 7085 7768: contig of 684 bp in length
 7769 7868: gap of 100 bp
 7869 8550: contig of 682 bp in length
 8551 8650: gap of 100 bp
 8651 9352: contig of 702 bp in length
 9353 9452: gap of 100 bp
 9453 10140: contig of 688 bp in length
 10141 10240: gap of 100 bp
 10241 10922: contig of 682 bp in length
 10923 11022: gap of 100 bp
 11023 11709: contig of 687 bp in length
 11710 11809: gap of 100 bp
 11810 12511: contig of 702 bp in length
 12512 12611: gap of 100 bp
 12612 13304: contig of 693 bp in length
 13305 13404: gap of 100 bp
 13405 14095: contig of 691 bp in length
 14096 14195: gap of 100 bp
 14196 14881: contig of 686 bp in length
 14882 14981: gap of 100 bp
 14982 15674: contig of 693 bp in length
 15675 15774: gap of 100 bp
 15775 16460: contig of 686 bp in length
 16461 16560: gap of 100 bp
 16561 17250: contig of 690 bp in length
 17251 17350: gap of 100 bp
 17351 18040: contig of 690 bp in length
 18041 18140: gap of 100 bp
 18141 18836: contig of 696 bp in length
 18837 18936: gap of 100 bp
 18937 19621: contig of 685 bp in length

* 19622 19721: gap of 100 bp
 19722 20425: contig of 704 bp in length
 20426 20525: gap of 100 bp
 20526 21224: contig of 699 bp in length
 21225 21324: gap of 100 bp
 21325 22017: contig of 693 bp in length
 22018 22117: gap of 100 bp
 22118 22815: contig of 698 bp in length
 22816 22915: gap of 100 bp
 22916 23609: contig of 694 bp in length
 23610 23709: gap of 100 bp
 23710 24402: contig of 693 bp in length
 24403 24502: gap of 100 bp
 24503 25191: contig of 689 bp in length
 25192 25291: gap of 100 bp
 25292 25981: contig of 690 bp in length
 25982 26081: gap of 100 bp
 26082 26774: contig of 693 bp in length
 26775 26874: gap of 100 bp
 26875 27563: contig of 689 bp in length
 27564 27663: gap of 100 bp
 27664 28355: contig of 692 bp in length
 28356 28455: gap of 100 bp
 28456 29145: contig of 690 bp in length
 29146 29245: gap of 100 bp
 29246 29943: contig of 698 bp in length
 29944 30043: gap of 100 bp
 30044 30745: contig of 702 bp in length
 30746 30845: gap of 100 bp
 30846 31535: contig of 690 bp in length
 31536 31635: gap of 100 bp
 31636 32329: contig of 694 bp in length
 32330 32429: gap of 100 bp
 32430 33114: contig of 685 bp in length
 33115 33214: gap of 100 bp
 33215 33898: contig of 684 bp in length
 33899 33998: gap of 100 bp
 33999 34694: contig of 696 bp in length
 34695 34794: gap of 100 bp
 34795 35486: contig of 692 bp in length
 35487 35586: gap of 100 bp
 35587 36292: contig of 706 bp in length
 36293 36392: gap of 100 bp
 36393 37095: contig of 703 bp in length
 37096 37195: gap of 100 bp
 37196 37872: contig of 677 bp in length
 37873 37972: gap of 100 bp
 37973 38663: contig of 691 bp in length
 38664 38763: gap of 100 bp
 38764 39449: contig of 686 bp in length
 39450 39549: gap of 100 bp
 39550 40236: contig of 687 bp in length
 40237 40336: gap of 100 bp
 40337 41025: contig of 689 bp in length
 41026 41125: gap of 100 bp
 41126 41818: contig of 693 bp in length
 41819 41918: gap of 100 bp
 41919 42608: contig of 690 bp in length
 42609 42708: gap of 100 bp
 42709 43399: contig of 691 bp in length
 43400 43499: gap of 100 bp
 43500 44169: contig of 670 bp in length
 44170 44269: gap of 100 bp
 44270 44966: contig of 697 bp in length
 44967 45066: gap of 100 bp
 45067 45750: contig of 684 bp in length
 45751 45850: gap of 100 bp
 45851 46545: contig of 695 bp in length
 46546 46645: gap of 100 bp
 46646 47336: contig of 691 bp in length
 47337 47436: gap of 100 bp
 47437 48123: contig of 687 bp in length
 48124 48223: gap of 100 bp

Alignment Scores:

US-09-763-712A-2_COPY_91_547 (1-457) x AC024368 (1-71044)

Search completed: March 21, 2003, 11:17:28
Job time : 3321.43 secs

